LEARNING IN EXTENSIVE-FORM GAMES, II EXPERIMENTATION AND NASH EQUILIBRIUM *

Drew Fudenberg
Department of Economics, Harvard University
and
David M. Kreps
Graduate School of Business, Stanford University, and
Berglass School of Economics, Tel Aviv University

June 1994

1. Introduction

The standard practice in economic applications of game theory is to assume that observed behavior in the situation being modeled will correspond to one of the Nash equilibria of the game. Yet Nash equilibrium supposes that all players have correct and independent beliefs about the off-path play of their opponents, and it is unclear why this should be the case. Several informal justifications have been suggested, among them the idea that players learn their opponents' strategies from repeatedly playing the game. However, as we argued in Fudenberg and Kreps (1994), general learning models need not lead to Nash equilibirum in extensive-form games: Repeated observations may lead players to learn the *actions* their opponents use along the equilibrium path of play, but players may not receive enough observations of their opponents' *off-path* play to justify the assumption that players know their opponents' *strategies*.

This paper investigates additional conditions under which learning does imply Nash equilibirum in general extensive-form stage games. We consider models in the general style of fictitious play: A small group of players interact repeatedly

*We are grateful to Robert Anderson, Robert Aumann, and David Levine for helpful comments. We would like to thank IDEI, Toulouse and the Institute for Advanced Studies, Tel-Aviv University, for their hospitality while this research was being conducted. The financial assistance of the National Science Foundation (Grants SES 88-08204, SES 90-08770, SES 89-08402, and SES 92-08954) and the John Simon Guggenheim Foundation is gratefully acknowledged.

in a single noncooperative game, called the stage game. The players' behavior is *asymptotically myopic*; i.e., players maximize (asymptotically) their immediate expected payoffs given their beliefs. Their beliefs are *asymptotically empirical*; i.e., if player i observes j acting in a particular situation a large number of times, i's prediction of j's behavior is close to the empirical frequencies with which j has acted in the past.

With these assumptions on beliefs and behavior, one can show that asymptotic steady states of the learning process must be Nash equilibria of the stage game, if the stage game is a game in strategic form and players learn, after each round of play, the strategy profile chosen by rivals (see Fudenberg and Kreps, 1993). But when the stage game is a game in extensive form and players observe only the actions taken by their rivals in the course of play, then any self-confirming equilibrium[1] is a candidate for an asymptotic steady state (see Fudenberg and Kreps, 1994). There are two reasons the set of self-confirming equilibria can contain non-Nash profiles. First, in a self-confirming equilibrium, two players can entertain diverse beliefs about the actions of a third, if the third moves at an information set that is off the path of play. Second, one player can entertain beliefs leading to correlations in the conjectured play by two or more rivals at information sets that are off the path of play. The learning models we have studied do not preclude either of these phenomena, because players need not learn about things they don't observe, and they may not observe behavior by their rivals at information sets that are off the path of the asymptotic steady state of the process.

If players experiment with suboptimal strategies/actions, however, it is possible that they obtain enough information about off-path play that only Nash equilibria are candidates for asymptotic steady states. In this paper, we build

[1] *Self-confirming equilibrium* is used in this paper as short-hand for Fudenberg and Levine's (1993a) *self-confirming equilibrium with unitary beliefs*.

directly on the analysis and results of Fudenberg and Kreps (1994), to see what it takes to obtain this type of result. Our analysis proceeds as follows.

We note first that, although Nash equilibrium supposes that players' beliefs are correct at every information set, observed play in a self-confirming equilibrium must be Nash if the beliefs meet the weaker requirement that each player's beliefs are correct at every information set that the player could cause to be reached. Intuitively, the player's optimal strategy depends only on his beliefs about play that are "relevant" in this sense; the player's optimal choice is not affected by beliefs about information sets the player cannot cause to occur.

To ensure that players' beliefs are correct at all of these relevant information sets, we develop assumptions that imply three things: all relevant information sets are reached infinitely often; beliefs at informations sets that are reached infinitely often converge to the empirical distribution of play at these information sets; and play at these information sets resembles the "target strategy" the stability of which we are testing. We formalize the latter two conditions by strengthening the asymptotic empiricism and asymptotic myopia conditions of our 1994 paper.

The condition that all relevant information sets are reached infinitely often proves more complicated to develop. We suppose that relevant information sets are reached because players *experiment* with suboptimal actions and/or strategies, and we formulate lower bounds on the frequency of experimentation that make it plausible that relevant information sets will be reached infinitely often.

Even if players do experiment "sufficiently frequently," some relevant information sets may not be reached infinitely often if, with high enough probability, two or more players experiment at the same dates. This problem is avoided if the players experiment at each date with probability bounded away from one, but that assumption is unappealing. We therefore develop an alternative assumption which serves the same purpose, based on the idea that players test the hypothe-

sis that the strategies used by their opponents are uncorrelated with the player's own choices. These statistical tests have the additional virtue of making our assumptions of asymptotic empiricism and asymptotic myopia more plausible.

Under these assumptions, we show that strategy profiles that are not Nash equilibria are unstable and that, conversely, all Nash profiles are weakly stable. Under somewhat weaker conditions on the players' off-path behavior, we show that outcomes (probability distributions over terminal nodes) are unstable if they cannot be generated by some Nash equilibrium. This result addresses the case where individual profiles fail to be stable only because the off-path play fails to converge, even if on-path behavior does converge. However the value of this result is questionable, as one of the assumptions required for it is highly suspect if we don't think off-path behavior is converging.

In Fudenberg and Kreps (1994), we argued that the Nash hypothesis of correct and independent beliefs at all information sets is unwarranted for off-path information sets, at least in the context of a learning story of our type. The results given here do provide conditions under which non-Nash steady states are impossible, but we believe that they should be viewed as primarily negative. The story we tell to rationalize Nash equilibrium calls for too much experimentation by the players and for behavior and beliefs rules that are too restrictive. If these assumptions are needed to guarantee that only Nash equilibria can be asymptotic steady states (begging the question of whether an asymptotic steady state will emerge at all), then assuming Nash in applications is too much.

Of course, there are possible rationales for Nash equilibrium other than the learning story considered here.[2] But on the basis of this analysis, our skepticism about the Nash hypothesis, applied to extensive form games, has increased. Our objective is to raise the reader's skepticism to a similar level.

[2] Examples include preplay communication and evolutionary processes such as the replicator dynamics.

2. *Preliminaries*

2.1. The stage game

We take as given some I-player extensive-form game, called the *stage game.* We imagine that the same I players play this game repeatedly, at dates $t = 1, 2, 3, \ldots$.

The stage game is a finite I-player extensive-form game of perfect recall. The set V is the set of nodes in the game tree, partially ordered by the precedence relation $\prec$; $Z \subseteq V$ is the subset of terminal nodes, and $X = V \setminus Z$ is the set of nodes where some player takes an action. The information sets $h \in H$ partition X; $h(x)$ is the information set containing x; $i(h)$ is the player who moves at h; H^i is the set of player i's information sets; and the information sets of i's opponents are denoted by $H^{-i} = H \setminus H^i$. The feasible actions at h are denoted $A(h)$ and are labeled so that $A(h) \cap A(h') = \emptyset$ for $h \neq h'$; $h(a)$ is the information set at which a is an available action. The set of feasible actions for player i is denoted A^i; the set of feasible actions for i's opponents is written A^{-i}. Because the game has perfect recall and each action $a \in A^i$ lies in a single $A(h)$, the precedence relation can be extended to $H^i \cup A^i \cup Z$; we write $h \prec h', a \prec h$, and so on, for $h, h' \in H^i$ and $a \in A^i$, for any player i. All of Nature's moves (if any) are placed at the start of the tree, so that each move by Nature corresponds to an initial node of the tree. The set of initial nodes is denoted W, with ϕ the objective initial probability distribution over W, which is known to all players. Player i's payoff if terminal node z is reached is $u^i(z)$; player i knows u^i. Our formal model is agnostic about whether players know the payoff functions of their opponents.

A pure strategy for player i in the stage game, written s^i, is a map from H^i to A^i satisfying $s^i(h) \in A(h)$; S^i is the set of all pure strategies for i. A mixed (behavior) strategy is a function π^i that maps each $h^i \in H^i$ to an element of the space $\Delta(A(h^i))$ of probability distributions over $A(h^i)$; Π^i denotes the space of

player i's behaviorally-mixed strategies. Pure and behaviorally-mixed strategy profiles are denoted by s and π, and are elements of $S = \prod_i S^i$ and $\Pi = \prod_i \Pi^i$, respectively. For pure and behaviorally-mixed strategy profiles for all players except i, we use symbols s^{-i} and π^{-i}, coming from the sets $S^{-i} = \prod_{j \neq i} S^j$ and $\Pi^{-i} = \prod_{j \neq i} \Pi^j$.

Each strategy profile π together with the initial probability distribution ϕ induces a probability distribution $\rho(\cdot|\pi)$ over the terminal nodes, which is computed under the assumption that each player's behavior is independent of the behavior of others. This probability distribution is called the *outcome* induced by π. In general, ρ will denote a probability distribution on Z, called an *outcome.*

Let $\bar{Z}(\pi)$ denote the support of $\rho(\cdot|\pi)$. Similarly, let $\bar{X}(\pi)$ be the set of all non-terminal nodes that have positive probability under π, and let $\bar{H}(\pi)$ denote the set of all information sets that π hits with positive probability.

2.2. *Histories*

Each play of the game at a given date results in a particular terminal node $z \in Z$ being reached, so the *history* at the beginning of round t of play is an element $\zeta_t = (z_1, \cdots, z_{t-1})$. (For $t = 1$, ζ_1 is used conventionally to denote the initial [informationless] history.) We assume that all players observe the outcome at the end of each round, so that all players know ζ_t at the start of round t. We use ζ to denote an infinite history of play $(z_1, z_2, \ldots)$, $\mathcal{Z}$ to denote the space of all infinite histories (so that $\mathcal{Z} = (Z)^\infty$), and $\mathcal{Z}_t$ to denote the space of all histories up time t (so that $\mathcal{Z}_t = (Z)^{t-1}$).[3]

Throughout this paper, we implicitly assume that all players know the game tree, the information sets, their own payoffs as a function of the terminal nodes, and ϕ. Each player assumes that his own behavior π^i and the behaviors π^j of each of his rivals are independent, so that if i plays according to π^i and is

[3] In general, subscripts will denote time and superscripts will denote players. The exceptional case of Z to the power $t-1$ is indicated by $(Z)^{t-1}$.

certain that rivals play according to π^{-i}, then the outcome of the game will be the terminal node z with probability $\rho(z|\pi)$.

We generally use $\kappa(\cdot;\zeta_t)$ as a counting function for the number of times the argument has "happened" up to time t, along history ζ_t. That is, $\kappa(v;\zeta_t)$ denotes the number of times node $v \in V$ was reached, $\kappa(h;\zeta_t)$ denotes the number of times information set $h \in H$ was reached, and $\kappa(a;\zeta_t)$ denotes the number of times action $a \in A$ was taken.

2.3. *Behavior rules*

Player i's *behavior rule* specifies the behavior strategy of i for each date t and each history ζ_t. We write $\hat{\pi}^i$ for a behavior rule of i, with $\hat{\pi}^i_t(\zeta_t)$ the behavior strategy for date t in history ζ_t. We write $\hat{\pi}^i_t(\zeta_t)(h)$ to refer to the behavior rule at information set $h \in H^i$, although we also write $\hat{\pi}^i_t(\zeta_t)(a)$ for the probability with which i takes a at t, given history ζ_t.

Behavior rules for each player, together with the probability distribution over initial nodes for each stage, give a probability distribution for the evolution of the entire system. We use $\mathbf{P}(\cdot)$ to denote this probability distribution, where the dependence on the players' behavior rules is implicit.

2.4. *Beliefs rules*

The players in our model will base their actions to some extent on predictions or assessments they make about the actions of their rivals. We model this with a *beliefs rule* $\hat{\gamma}^i$ for each player i, which gives for each date t and partial history ζ_t a probability distribution $\gamma^i_t(\zeta_t)$ on the (behavioral) strategy profile of i's rivals. That is, $\gamma^i_t(\zeta_t)$ is a probability distribution on Π^{-i}. We interpret this as i's beliefs as to the profile his rivals are about to use. Note that $\pi^{-i} \in \Pi^{-i}$ involves independent play by i's rivals, but $\hat{\gamma}^i_t(\zeta_t)$, being a mixture of π^{-i}, allows i to have correlated conjectures about his rivals' play.

For given t and ζ_t, i's immediate expected payoff if he plays strategy π^i,

relative to his beliefs, is

$$u^i(\pi^i, \hat{\gamma}^i_t(\zeta_t)) = \int_{\pi^{-i} \in \Pi^{-i}} \sum_{z \in Z} u^i(z) \rho(z|\pi^i, \pi^{-i}) \hat{\gamma}^i_t(\zeta_t)(\mathrm{d}\pi^{-i}).$$

2.5. *Review from Fudenberg and Kreps (1994)*

We conclude this section by reviewing the central definitions and results of Fudenberg and Kreps (1994).

For all $\zeta \in \mathcal{Z}$, let $H_{\text{p.f.}}(\zeta)$ be those information sets that are reached a strictly positive fraction of the time along the history ζ, using a limit infimum test; i.e., $h \in H_{\text{p.f.}}(\zeta)$ if $\liminf_{t \to \infty} \kappa(h; \zeta_t)/t > 0$.

Definition. *Player i's belief rule $\hat{\gamma}^i$ is* ***asymptotically empirical*** *if for every $\epsilon > 0$, every infinite history ζ, every information set $h^j \in H_{\text{p.f.}}(\zeta) \cap H^j$ for $j \neq i$, and every $a \in A(h^j)$,*

$$\lim_{t \to \infty} \hat{\gamma}^i_t(\zeta_t)\left(\left\{\pi^{-i} \ : \ \left|\pi^j(a) - \frac{\kappa(a; \zeta_t)}{\kappa(h^j; \zeta_t)}\right| < \epsilon\right\}\right) = 1. \tag{2.1}$$

In words, i's beliefs as to what will happen at information set h^j must be converging to a neighborhood of the empirical record of actions taken at previous visits to h^j if h^j has been visited a nonvanishing fraction of the time.

Definition. *Fix player i and i's beliefs rule $\hat{\gamma}^i$. The behavior rule $\hat{\pi}^i$ for i is* ***asymptotically myopic with calendar-time limitations on experimentation*** *if there exist: (1) a sequence strictly positive numbers $\{\epsilon_t\}$ with $\lim_{t \to \infty} \epsilon_t = 0$, (2) a nondecreasing sequence of nonnegative integers $\{\eta_t; t = 1, 2, \ldots\}$ with $\lim_{t \to \infty} \eta_t/t = 0$; (3) behavior rules $\check{\pi}^i$ and $\bar{\pi}^i$ for i; and (4) for each t, ζ_t, and $h \in H^i$, a number $\check{\alpha}^i_t(\zeta_t)(h) \in [0, 1]$, such that:*

(a) For all t, ζ_t, and $h \in H^i$, $\hat{\pi}^i_t(\zeta_t)(h) = \check{\alpha}^i_t(\zeta_t)(h) \times \check{\pi}^i_t(\zeta_t)(h) + (1 - \check{\alpha}^i_t(\zeta_t)(h)) \times \bar{\pi}^i_t(\zeta_t)(h)$.

(b) For all t, ζ_t, *and* $h \in H^i$, $u^i(\breve{\pi}_t^i(\zeta_t), \hat{\gamma}^i(\zeta_t)) + \epsilon_t \geq max_{s^i \in S^i} u^i(s^i, \hat{\gamma}^i(\zeta_t))$.

(c) If $\breve{\alpha}_t^i(\zeta_t)(h) < 1$, *then* $\kappa(a'; \zeta_t) \leq \eta_t$ *for some* $a' \in A(h)$, *and* $\tilde{\pi}_t^i(\zeta_t)(h)$ *gives positive probability only to actions* $a \in A(h)$ *such that* $\kappa(a; \zeta_t) \leq \eta_t$.

We call $\breve{\pi}^i$ the *nonexperimental* portion of i's behavior rule, with $\tilde{\pi}^i$ the *experimental* portion. The interpretation is that player i decides, information set by information set, whether to experiment, with $\breve{\alpha}_t^i(\zeta_t)(h)$ the probability that i doesn't experiment at h. Experiments are only permitted with actions that have been taken infrequently relative to calendar time (part (c)), and the nonexperimental portion of i's strategy must be asymptotically myopically optimal, on an ex ante basis (part (b)).[4]

In the following definitions the term *model* is used to mean a specification of beliefs rules and behavior rules, one each for each player in the stage game. Recall that $\mathbf{P}(\cdot)$ denotes the probability distribution over $\mathcal{Z}$ induced by any fixed model and the (fixed) initial distribution over initial nodes in the stage game.

Definitions.[5] *(a) A strategy profile* π_* *is* ***unstable*** *relative to a given class of models if there exists* $\epsilon > 0$ *such that, for every model from this class,*

$$\mathbf{P}\big(\|\hat{\pi}_t(\zeta_t) - \pi_*\| < \epsilon \text{ for all } t\big) = 0.$$

(b) The outcome ρ_* *is* ***unstable*** *relative to a given class of models if there exists* $\epsilon > 0$ *such that, for every model from this class,*

$$\mathbf{P}\big(\|\rho(\hat{\pi}_t(\zeta_t)) - \rho_*\| < \epsilon \text{ for all } t\big) = 0.$$

---

[4] Part (b) of the definition prevents $\breve{\pi}$ from assigning a large probability to an "experimental" action, but since $\breve{\pi}$ can be slightly (vanishingly) suboptimal, $\breve{\pi}$ can assign (vanishingly) small probability to suboptimal actions. Thus while deterministic experiments cannot be accomodated within $\breve{\pi}^i$, "experiments taken at random" can be; cf. Fudenberg and Kreps (1994).

[5] Throughout, $\hat{\pi}_t(\zeta_t)$ is the profile of strategies employed by the players at date t with history ζ_t, and $\|\cdot\|$ denotes the sup norm in whatever (finite-dimensional Euclidean) space is appropriate.

(c) The profile π_* *is* ***locally stable*** *relative to a given class of models if for some model out of this class,*

$$\mathbf{P}\big(\lim_{t\to\infty} \pi_t(\zeta_t) = \pi_*\big) > 0.$$

Definition. *The strategy profile* π_* *is a* ***self-confirming equilibrium*** [6] *if for each player* i *there are beliefs* γ_*^i *such that*

(a) π_*^i *maximizes* $u^i(\pi^i, \gamma_*^i)$, *and*

(b) $\gamma_*^i\big(\{\pi^{-i} : \pi^j(h^j) = \pi_*^j(h^j) \text{ for all } j \neq i \text{ and } h^j \in \bar{H}(\pi_*)\}\big) = 1.$

Proposition 2.1. *For the class of models where beliefs rules are all asymptotically empirical and behavior rules are all asymptotically myopic with calendar-time limitations on experiments: (a) Strategy profiles that are not self-confirming equilibrium profiles are unstable. (b) Moreover, outcomes that are not the outcomes arising from a self-confirming equilibrium are unstable. (c) Every self-confirming equilibrium strategy profile is locally stable.*

3. *Relevant information sets*

A self-confirming equilibria is not Nash because of incorrect assessments by some players about what happens off the path of play. But to know that a self-confirming equilibrium strategy profile is a Nash equilibrium, we *do not* need to assume that each player has correct assessments at *every* information set. Any given player can be wrong about what would happen at information sets that cannot be reached unless others deviate. It is sufficient that assessments are correct at information sets that are relevant in the following sense.

Definition. *An information set* $h \in H$ *is* ***relevant to player*** i ***at the profile*** π_* *if* $h \in \bar{H}(\pi^i, \pi_*^{-i})$ *for some* $\pi^i \in \Pi^i$.

[6] In terms of the original definition in Fudenberg and Levine (1993a), this is a self-confirming equilibrium *with unitary beliefs*.

This definition is phrased "objectively"; it speaks about information sets that are relevant to player i given a profile of behavior by i and his rivals. An alternative, "subjective" definition would fix i's beliefs at γ^i and ask which information sets i *believes* might be reached, as he changes his own strategy.

For the set of information sets relevant to i at the profile π_*, we write

$$\hat{H}^i(\pi_*) = \bigcup_{\pi^i \in \Pi^i} \bar{H}(\pi^i, \pi_*^{-i}).$$

We will also write $\hat{H}^i(\pi_*^{-i})$ for information sets relevant to i given a profile of strategies for his rivals.

To illustrate this definition, fix a strategy profile π_*. For each player i, all the information sets that lie along the path of play are relevant to i; i.e., $\bar{H}(\pi_*) \subseteq \hat{H}^i(\pi_*)$ for all i. If player i is never called upon to move at the strategy profile π_* (i.e., if $H^i \cap \bar{H}(\pi_*) = \emptyset$), then $\bar{H}(\pi_*) = \hat{H}^i(\pi_*)$. If player i does move along the path of play at π_*, all information sets that he can cause to happen (with positive probability) by deviating along the path are relevant, as are information sets that he can cause to happen by a further deviation at an information set he causes to happen by a deviation along the path of play, and so on.

Proposition 3.1. *If a strategy profile π_* and beliefs $(\gamma_*^1, \ldots, \gamma_*^I)$ satisfy*

(a) for each i, π_*^i is a best response against the beliefs γ_*^i; i.e., $u^i(\pi_*^i, \gamma_*^i) = \max_{s^i \in S^i} u^i(s^i, \gamma_*^i)$; and

(b) for each player i,

$$\gamma_*^i\left(\{\pi^{-i} \;:\; \pi^{-i}(h) = \pi_*^{-i}(h) \text{ for all } h \in \hat{H}^i(\pi_*^{-i}) \setminus H^i\}\right) = 1,$$

then π_* is a Nash equilibrium.

In words, given a strategy profile and beliefs for each player that rationalize the player's part of the strategy profile, the profile is a Nash equilibrium if each

player's beliefs are correct at information sets that are relevant to the player.[7] Compare with the definition of a self-confirming equilibrium, where beliefs are necessarily correct only at information sets along the path of play.

The proof of Proposition 3.1 is a straightforward corollary to the following lemma, which (in turn) is a matter of stringing definitions together.

Lemma 3.1. *Fix* π^{-i}, *and let* γ^i *be any set of beliefs for player* i *such that*

$$\gamma^i\big(\{\tilde{\pi}^{-i} \;:\; \tilde{\pi}^{-i}(h) = \pi^{-i}(h) \text{ for all } h \in \hat{H}^i(\pi^{-i}) \setminus H^i\}\big) = 1.$$

Then for all π^i, $u^i(\pi^i, \pi^{-i}) = u^i(\pi^i, \gamma^i)$.

The following is a useful "contrapositive" to Proposition 3.1.

Proposition 3.2. *If a strategy profile* π_* *is not a Nash equilibrium, there there exists an* $\epsilon' > 0$, *a player* i, *and a strategy* $\tilde{\pi}^i \in \Pi^i$ *such that for all beliefs* γ^i *satisfying*

$$\gamma^i\Big(\big\{\pi^{-i} \;:\; \max_{j \neq i, h^j \in \hat{H}^i(\pi_*^{-i})} \|\pi^j(h^j) - \pi_*^k(h^j)\| < \epsilon'\big\}\Big) > 1 - \epsilon',$$

and for all π^i *such that* $\|\pi^i - \pi_*^i\| < \epsilon'$,

$$u^i(\tilde{\pi}^i, \gamma^i) > u^i(\pi^i, \gamma^i) + \epsilon'.$$

Proof. Although this is fairly standard (following Lemma 3.1), we provide some of the details for completeness. Since π_* is not a Nash equilibrium, there exists a player i and a strategy $\tilde{\pi}^i$ for player i such that

$$u^i(\tilde{\pi}^i, \pi_*^{-i}) > u^i(\pi_*^i, \pi_*^{-i}). \tag{3.1}$$

[7] For at least two reasons, this condition is sufficient but not necessary. First, for some payoff functions, much less information may be required. For example, if π_*^i is a strict best response to γ_*^i for each i, beliefs can be slightly incorrect everywhere and we have a Nash equilibrium profile. Taking this to an extreme, if each player has a strictly dominant strategy, it doesn't matter what each believes. Second, even for general payoff functions, our definition of relevant information sets is too inclusive. If player i doesn't move along the path of π_*, then i's beliefs are irrelevant. More generally, if $\rho(s^i, \pi_*^{-i})(z)$ is unaffected by s^i for z that follow some $h \in H^j$, then i's beliefs about what happens at h are irrelevant.

Suppose that the conclusion of the proposition is false. Then for every n we can find some π_n^i and γ_n^i such that

$$\gamma_n^i\Big(\big\{\pi^{-i} : \max_{j\neq i, h^j \in \hat{H}^i(\pi_*^{-i})} \|\pi^j(h^j) - \pi_*^k(h^j)\| < 1/n\big\}\Big) > (n-1)/n, \tag{3.2}$$

$\|\pi_n^i - \pi_*^i\| < 1/n$, and yet

$$u^i(\pi_n^i, \gamma_n^i) + 1/n \geq u^i(\tilde{\pi}^i, \gamma_n^i). \tag{3.3}$$

(In this proof, subscripts n refer to a sequence and not to the usual index of time.) Each Π^j is a compact subset of a complete, separable metric space, so the space of beliefs by player i (the space of probability measures on $\prod_{j\neq i} \Pi^j$) is a compact subset of a complete separable metric space under the topology of weak convergence. We can therefore find beliefs γ_*^i such that, along a subsequence, γ_n^i converges to γ_*^i in the weak topology. Of course, π_n^i converges to π_*^i along the sequence. Since u^i is continuous in both its arguments (the second, in the weak topology), passing to the limit along this subsequence in (3.3) yields

$$u^i(\pi_*^i, \gamma_*^i) \geq u^i(\tilde{\pi}^i, \gamma_*^i). \tag{3.4}$$

And passing to the limit in (3.2) yields

$$\gamma_*^i\Big(\big\{\pi^{-i} : \pi^j(h^j) = \pi_*^k(h^j) \text{ for all } h^j \in \hat{H}^i(\pi_*^{-i})\big\}\Big) = 1. \tag{3.5}$$

But then by Lemma 3.1, an implication of (3.5) is that

$$u^i(\pi_*^i, \gamma_*^i) = u^i(\pi_*^i, \pi_*^{-i}) \text{ and } u^i(\tilde{\pi}^i, \gamma_*^i) = u^i(\tilde{\pi}^i, \pi_*^{-i}).$$

Compare this with (3.1) and (3.4), and a contradiction is obtained. ∎

4. *Beliefs and behavior at off-path information sets*

In order to conclude that players learn enough about behavior at relevant off-path information sets, we will make assumptions that ensure that all relevant information sets are reached infinitely often. But this is insufficient to eliminate non-Nash profiles, given our definitions of asymptotic empiricism and asymptotic myopia. Neither of these conditions imposes restrictions on beliefs or behavior at information sets reached a vanishing fraction of the time, even if the information set is reached infinitely often. We therefore strengthen each.

Recall that $H_{\text{p.f.}}(\zeta)$ is the set of information sets reached a nonvanishing fraction of the time along the history ζ. In a similar spirit, let $H_{\text{i.o.}}(\zeta)$ be the collection of information sets reached infinitely often along ζ; i.e.,

$$H_{\text{i.o.}}(\zeta) = \{h \in H : \lim_{t\to\infty} \kappa(h;\zeta_t) = \infty\}.$$

Definition. *Player i's belief rule $\hat{\gamma}^i$ is* ***strongly asymptotically empirical*** *if for every $\epsilon > 0$, every infinite history ζ, every information set $h^j \in H_{\text{i.o.}}(\zeta) \cap H^j$ for $j \neq i$, and every $a \in A(h^j)$,*

$$\lim_{t\to\infty} \hat{\gamma}^i_t(\zeta_t)\left(\left\{\pi^{-i} \;:\; \left|\pi^j(a) - \frac{\kappa(a;\zeta_t)}{\kappa(h^j;\zeta_t)}\right| < \epsilon\right\}\right) = 1. \tag{4.1}$$

Definition. *Fix player i and i's beliefs rule $\hat{\gamma}^i$. The behavior rule $\hat{\pi}^i$ for i is* ***asymptotically myopic with experience-time limitations on experimentation*** *if there exist: (1) a sequence strictly positive numbers $\{\epsilon_t\}$ with $\lim_{t\to\infty} \epsilon_t = 0$, (2) a nondecreasing sequence of nonnegative integers $\{\eta_t; t = 1,2,\ldots\}$ with $\lim_{t\to\infty} \eta_t/t = 0$; (3) behavior rules $\check{\pi}^i$ and $\tilde{\pi}^i$ for i; and (4) for each t, ζ_t, and $h \in H^i$, a number $\check{\alpha}^i_t(\zeta_t)(h) \in [0,1]$, such that:*

(a) For all t, ζ_t, and $h \in H^i$, $\hat{\pi}^i_t(\zeta_t)(h) = \check{\alpha}^i_t(\zeta_t)(h) \times \check{\pi}^i_t(\zeta_t)(h) + (1 - \check{\alpha}^i_t(\zeta_t)(h)) \times \tilde{\pi}^i_t(\zeta_t)(h)$.

(b)

(c) If $\check{\alpha}_t^i(\zeta_t)(h) < 1$, *then* $\kappa(a';\zeta_t) \leq \eta_{\kappa(h;\zeta_t)}$ *for some* $a' \in A(h)$, *and* $\tilde{\pi}_t^i(\zeta_t)(h)$ *gives positive probability only to actions* $a \in A(h)$ *such that* $\kappa(a;\zeta_t) \leq \eta_{\kappa(h;\zeta_t)}$.

In words, strong asymptotic empiricism requires that players' inferences converge to the empirical distribution at all information sets reached infinitely often, and asymptotic myopia with experience-time limitations requires that experimentation at any information set takes place a vanishing fraction of the time that the information set is visited.

While these two strengthened assumptions are formally simple, it is less easy to see how reasonable they are as behavioral assumptions, at least compared to the weaker restrictions from Fudenberg and Kreps (1994).

Concerning strong asymptotic empiricism, note first that if player i believes his rivals are playing some (unknown to him) strategy profile π^{-i}, he computes posterior assessments concerning π^{-i} after each round using Bayes' rule, and his prior on π^{-i} is sufficiently diffuse (nondoctrinaire), then his beliefs rule will be strongly asymptotically empirical. A similar conclusion holds if i believes that behavior by any rival at any information set h will settle down to repeated play of some fixed probability distribution over $A(h)$, when and if the number of visits to h approaches infinity.

But as discussed in Fudenberg and Kreps (1994, Section 4), players may be justifiably skeptical about the empirical record of behavior at information sets that are visited with vanishing frequency, at least insofar as players are assumed to be (only) asymptotically myopic: If j believes that there is vanishing probability that $h \in H^j$ can be reached by any strategy she might attempt, then her actions at h have vanishing impact on her expected payoffs, and thus her nonexperimental play at h can be erratic without violating asymptotic myopia. Insofar as j's beliefs about the relevance of h correspond to increasingly infrequent visits to

h [8], and insofar as i is aware of this, i may be reluctant to trust to empirical evidence of how j acts there. Moreover, if experimentation is limited by calendar time, then any actions can be accomodated as experiments at information sets visited with (sufficiently rapidly) vanishing frequency.

Our point is simple: Asymptotic empiricism applied at any information set requires the presumption that behavior there is asymptotically i.i.d. This is a strong presumption for players to make about information sets visited a non-vanishing fraction of time, but we find it even more heroic for information sets visited infinitely often but a vanishing fraction of time. [9]

We also have a hard time justifying experience-time limitations on experimentation. Because we have avoided anything like a precise value of information story, a formal justification is infeasible. And informal justifications that we have concocted are somewhat tortured and (we believe) not altogether convincing. In fact, experience-time limitations on experimentation are necessary for some of our results (showing that non-Nash strategy profiles are unstable) but not for others (instability of non-Nash outcomes). We find it expositionally easiest to discuss this assumption and the role it plays in our analysis after conducting that analysis, so we defer further discussion until Section 8.

5. *Reaching relevant information sets infinitely often*

We now study assumptions on behavior that "nearly" imply that every information set relevant at π_* is reached infinitely often with probability one, if

[8] This is not a direct implication, of course.

[9] Note in this regard that the presumption is a degree more palatable in a setting with anonymous random matching within a large population that regenerates (through, say, a process of birth and death). It is possible that participants in the large population act in a fashion that is correlated with the calendar date, but we find it more plausible in such a setting that players would regard the actions of their anonymous rivals as draws from an (asymptotically) i.i.d. distribution, which in turn would justify strong asymptotic empiricism.

nonexperimental behavior $\check{\pi}_t(\zeta_t)$ remains within some small-enough neighborhood of π_*.[10] These assumptions put lower bounds on the rate and/or number of experiments that players undertake.

These assumptions can be developed in (at least) two different ways. One can think in terms of experimentation information set by information set; i.e., players experiment with *actions* at each of their information sets, without (explicitly) considering how those actions string together into experiments with *strategies*. Or one can think about experimentation at the strategic level: At date t, player i chooses a strategy s^i to play at that date (possibly according to some mixed strategy distribution), so that if the player is going to experiment (play a distinctly suboptimal pure strategy), the experiment is formed in terms of behavior at all information sets in H^i.

The two different approaches lead to the same basic conclusions, but they require somewhat different notation and assumptions. It is expositionally the most easy for us to fix on one approach, carry it through, and then return to the second approach, rather than carry both forward at once. Accordingly, we will work with experimentation information set by information set; in an appendix, we discuss (without all the details) the other approach.

5.1. Minimal experimentation at a single information set

Definition. *For a given player i and information set $h \in H^i$, the behavior rule $\hat{\pi}^i$ satisfies the* ***minimal experience-time experimentation condition at*** *h if there exists a strictly positive constant β and a nondecreasing sequence of nonnegative integers $\{\nu_k\}$ satisfying $\nu_k \to \infty$ and $\nu_k/k \to 0$ such that, for all t and ζ_t, if*

$$\kappa(a;\zeta_t) < \nu_{\kappa(h(a);\zeta_t)} \tag{5.1}$$

[10] The meaning of "nearly" will become clear as we proceed.

for at least one action $a \in A(h)$, *then*

$$\hat{\pi}_t^i(\zeta_t)(\{a \in A(h) \ : \ \kappa(a;\zeta_t) \leq \nu_{\kappa(h(a);\zeta_t)}\}) \geq \beta. \tag{5.2}$$

In words, there is a lower bound on the probability of taking those actions at h that have been taken rarely relative to the number of opportunities to act at h.

It should be clear that the requirements of this condition can be made consistent with experience-time limitations on experimentation: As long as $\nu_k \leq \eta_k$ (where ν_k comes from this definition and η_k from the asymptotic myopia condition), players are permitted by the earlier definition to take the experiments that are required here.

The force of this condition is most easily seen in the following result.

Proposition 5.1. *Fix behavior rules for all the players. If* $\hat{\pi}^i$ *satisfies the minimal experience-time experimentation condition at* h, *then for every* $a \in A(h)$,

$$\mathbf{P}\left(\left\{\zeta \ : \ \lim_{t\to\infty} \kappa(h;\zeta_t) = \infty \text{ and } \lim_{t\to\infty} \kappa(a;\zeta_t) < \infty\right\}\right) = 0.$$

As a rough paraphrase, if h is visited infinitely often, then every action available at h must be taken infinitely often.

Proof. Fix h. For each ζ, let ι_k be a random variable equal to 1 if, on the kth visit to h, an action a is taken for which (5.1) holds, and $\iota_k = 0$ if some other action is taken. (If h is not reached k times, $\iota_k = 0$.) For each $a \in A(h)$, let $\iota_k(a) = 1$ if $\iota_k = 1$ and the action taken at the kth visit to h is a. Note that at $\iota_k(a) = 1$ for at most one a, for each k.

Suppose that for h and some specific a^*,

$$\mathbf{P}\left(\left\{\zeta \ : \ \lim_{t\to\infty} \kappa(h;\zeta_t) = \infty \text{ and } \lim_{t\to\infty} \kappa(a^*;\zeta_t) < \infty\right\}\right) > 0.$$

Then we know that for sufficiently large K,

$$\mathbf{P}\left(\left\{\zeta \;:\; \lim_{t\to\infty} \kappa(h;\zeta_t) = \infty \text{ and } \lim_{t\to\infty} \kappa(a^*;\zeta_t) < K\right\}\right) > 0. \tag{5.3}$$

But then, on the event defined in (5.3), when t becomes sufficiently large, (5.1) holds for a^*, and therefore on this event there is probability at least β that $\iota_k = 1$ for all sufficiently large k. Thus on the event of positive probability in (5.3), $\liminf_{k\to\infty}\left[\sum_{j=1}^{k} \iota_j\right]/k \geq \beta$ almost surely. The intuition should now be clear: Actions taken because (5.1) applies will be taken a nonvanishing frequency of the time (almost surely). It cannot be, therefore, that each occurs no more than a vanishing frequency of the time.

To formalize this intuition, let M be the number of actions in $A(h)$. We claim that

$$\sum_{j=1}^{k} \iota_j \leq M(\nu_k + 1). \tag{5.4}$$

This estimate will complete the proof, since by assumption, $\nu_k/k \to 0$, and hence $M(\nu_k + 1)/k \to 0$.

To derive the estimate (5.4), write

$$\sum_{j=1}^{k} \iota_j = \sum_{a\in A(h)} \sum_{j=1}^{k} \iota_j(a).$$

Fix any action a, and consider $\sum_{j=1}^{k} \iota_j(a)$. Let j' be the largest index from 1 to k such that $\iota_j(a) = 1$. (If $\iota_j(a) = 0$ for all j between 1 and k, there is nothing to prove.) Then we know that $\sum_{j=1}^{k} \iota_j(a) = \sum_{j=1}^{j'} \iota_j(a)$. But by the definition of ι, at the j' th visit to h, a had been taken no more than $\nu_{j'}$ times, thus $\sum_{j=1}^{j'} \iota_j(a) \leq \nu_{j'} + 1$. Since ν_j is nondecreasing, we have the desired estimate. ∎

This result indicates both the mathematical strength and the intuitive unpalatability of mimimal experience-time experimentation. In order to ensure that

every relevant information set is reached infinitely often, we must assume that players experiment infinitely often with suboptimal actions. The assumption does just that. But *any* assumption that implies a positive probability of an infinite number of sub-optimal experiments will be inconsistent with optimal behavior in the discounted multi-armed bandit problem, where (with a full-support prior) the optimal solution involves a halt to experimentation in finite time with probability one, with positive probability of "locking on" to the objectively "wrong" arm. Any assumption that causes relevant off-path information sets to be reached infinitely often is qualitatively wrong as a description of optimal behavior for a discounted-expected-utility maximizer in a bandit problem.

However, as the player's discount factor increases towards one, the value of information increases, so the player tends to take more experiments, and the probability of locking onto the wrong arm goes to zero. Our assumptions involving an infinite number of experiments can be viewed as an approximation to optimal behavior in the bandit problem for discount factors close to one; a key to evaluating whether conditions requiring infinite experimentation give unrealistic conclusions for expected-utility-maximizing players is the degree of player impatience. [11]

Moreover, optimal behavior for a player seeking to maximize time-average expected payoffs, on the other hand, necessarily involves infinite experimentation with every arm, albeit at a vanishing rate (for subjectively suboptimal arms). Indeed, any strategy (in a classic multi-armed bandit problem) that calls for infinite experimentation with each arm but vanishingly frequent play of subjectively suboptimal arms will, almost surely, maximize time-average expected payoffs.

A second intuitive flaw in the definition arises because the definition ensures that *every* action at h will be taken infinitely often if h is reached infinitely often.

[11] For a precise analysis of these matters in a context a bit different from the one here, see Fudenberg and Levine (1993b).

Suppose that a player did wish to visit every relevant information set infinitely often, in order to learn how his rivals act. In some cases, there might be several actions that reach a rival's information set and that have different immediate expected costs. In such cases, it makes sense to suppose that the player would choose to experiment with the cheapest action.

Accordingly, the minimal experience-time experimentation condition, as defined, is too restrictive on reasaonble behavior. In order to keep the level of technical difficulties manageable, we will live with this over-restrictiveness in this paper; the definition will not be modified to accomodate this sort of consideration. But we note in passing that this consideration can play an important role when it comes to studying equilibrium refinements; it leads to refinements with the flavor of Myerson's (1978) properness criterion, because out-of-equilibrium actions, which are experiments, tend to be taken as cheaply as possible.

5.2. The MME condition

The definition of minimal experience-time experimentation is made at a single information set h. For a general extensive-form game, H^i will consist of more than a single information set, and so we must now ask, At which information sets in H^i should i be required to experiment? The simplest answer one can imagine is to insist that the condition hold for all $h \in H^i$. However this is a stronger assumption than is required for our results, and it is also unpalatable intuitively.

Consider, for example, the game of complete and perfect information shown in Figure 1. Imagine that along some history of play ζ, player 1 chooses A_1 a vanishing frequency of the time, and that player 2 begins with beliefs about the actions of player 3 that lead to the *ex ante* assessment that 3 is at least as likely to choose D_3 as A_3. Then player 2, faced with a choice between A_2 and D_2 (given the opportunity to move) would see A_2 as representing a costly experiment, but one that (potentially) could lead to valuable information. However for that

information to be valuable, two things must happen. Player 2 must discover that player 3 will choose A_3 with high probability, and player 1 must give player 2 the opportunity to use that information by choosing A_1. If player 2 is sufficiently patient, she might be willing to experiment with A_2, in order to test the first part of this two-part condition, but only as long as she has reason to believe that the second part holds (which has nothing to do with her choice whether to experiment). And if player 1 chooses A_1 a vanishing frequency of the time, then player 2 will conclude that she will not get many opportunities to use the information, which might cause her to abandon experimentation with A_2. Even if she is very patient, as with a per-period disount factor close to one, she will not see much future gain in getting the information if she believes it will be many rounds before she gets the opportunity to use that information.

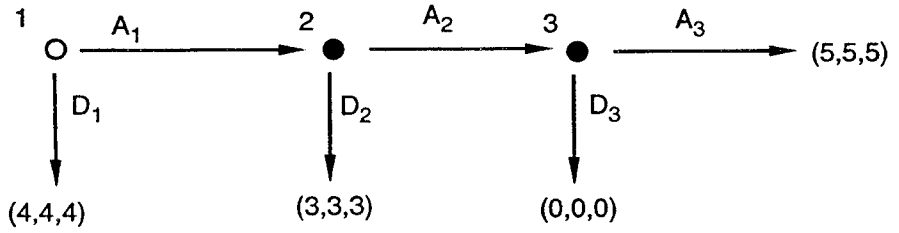

FIG. 1. A stage game.

N.B., if player 2 believes that player 1 is unlikely to choose A_1, she suffers a very small *ex ante* loss from choosing A_2. But we are not saying that player 2 necessarily will abandon experimentation at her information set (if player 1 chooses A_1 a vanishing fraction of the time), only that it is unreasonable to insist that she continue to experiment.

Accordingly, we seek a condition that mandates a minimal level of experimentation only at information sets that seem relevant to the player.

To make the formal definition, we consider player i's decision tree, which consists of $H^i \cup A^i \cup Z$, with precedence inherited from $\prec$ in the usual fashion

(because of perfect recall). For any $h \in H^i$ that is not an initial information set in i's decision tree, we write $a_p^i(h)$ for h's immediate predecessor in the decision tree; i.e., $a_p^i(h) = a'$ where a' is the last action i takes prior to h that is necessary if h is going to be reached. Also, we write $\Phi(h)$ for all information sets $h' \in H^i$ that precede h together with h itself. To make the definition notationally neat, if h is an initial information set for player i, we use the convention $\kappa(a_p^i(h);\zeta_t) = t-1$, even though, strictly speaking, there is no $a_p^i(h)$.

Definition. *The behavior rule $\hat{\pi}^i$ of player i satisfies the* ***modified minimum experience-time experimentation condition*** *(hereafter, the* ***MME*** *condition), if there exists a strictly positive constant β, a nondecreasing sequence of nonnegative integers $\{\nu_k\}$ satisfying $\nu_k \to \infty$ and $\nu_k/k \to 0$, and a nonincreasing sequence of strictly positive numbers $\{\delta_k\}$ satisfying $\delta_k \to 0$, such that, for all t and ζ_t, if*

(a) $h \in H^i$ satisfies $\kappa(h';\zeta_t)/\kappa(a_p^i(h');\zeta_t) \geq \delta_{\kappa(a_p^i(h');\zeta_t)}$ for all $h' \in \Phi(h)$, and

(b) $\kappa(a;\zeta_t) < \nu_{\kappa(h(a);\zeta_t)}$ for at least one action $a \in A(h)$,

then $\hat{\pi}_t^i(\zeta_t)(\{a \in A(h) : \kappa(a;\zeta_t) < \nu_{\kappa(h(a);\zeta_t)}\}) \geq \beta$.

In words, the requirement to try actions that have been taken rarely (relative to the opportunities to do so) is limited to information sets h that, at the given time and history, satisfy condition (a). If condition (a) fails, then experiments are not required. The idea behind condition (a) is that i needn't experiment at information set $h \in H^i$ if for some $h' \in \Phi(h)$, the data suggest that there is vanishingly small probability that i's rivals will cause the transition from $a_p^i(h')$ to h'. In such a case, i is released from the need to experiment at h (and moreover at h' and all its successors) because, presumably, there may be (asymptotically) zero value in the information. [12]

[12] N.B., i is released from the *requirement* to experiment at h; for situations as in Figure 1(a), he may still choose to do so. Nothing we do here precludes him from making this choice.

To see the impact of the MME condition vs. requiring minimal experience-time experimentation at every information set, consider games of complete and perfect information. It is relatively easy to prove that if each player's behavior rule satisfies the minimal experience-time experimentation condition at every information set, then in any game with complete and perfect information, every information set will be reached infinitely often, with probability one. Thus if we sharpen asymptotic myopia by requiring ex-post calculations of expected payoffs,[13] we can prove that *In any game of complete and perfect information, if each player's behavior rule satisfies the minimal experience-time experimentation condition at every information set, and if players use ex-post payoff calculations in the definition of asymptotic myopia, then any strategy profile that isn't a subgame perfect Nash equilibrium is unstable.*

Compare this with MME and (in particular) with the game in Figure 1. Of course, player 1 must experiment with A_1 infinitely often, even under MME. If player 2 experiments with A_2 infinitely often (and if player 3 used ex post evaluation), this would lead player 2 eventually to choose A_2 with frequency approaching one, leading player 1 to choose A_1 with frequency approaching one. (That is, A_1, A_2, A_3 is the unique subgame perfect Nash equilibrium.) But if player 1 believes that player 2 will choose D_2 with high probability, player 1 believes that D_1 is his short-run optimal choice. Thus (as long as nothing changes player 1's mind about 2's strategy) player 1 will choose A_1 a vanishing fraction of the time. Under MME, this can lead player 2 to abandon experimentation with A_2. Thus under MME, the subgame-imperfect Nash equilibrium D_1, D_2, A_3 will be stable, even assuming the players use ex post evaluation.

This rationale for condition (a) suggests that the need to experiment should be tested using i's beliefs and not empirical frequencies. That is, given i's

[13] To be precise, require that $\breve{\pi}_t^i(\zeta_t)$ is no worse than ϵ_t -suboptimal in terms of i's beliefs about the actions of his rivals, computed for and conditional upon reaching each of i's information sets, where we restrict each player to full-support belief rules for every finite date t and history ζ_t.

beliefs rule $\hat{\gamma}^i$, we can construct i's conditional assessment that a transition will be made from $a_p^i(h')$ to h', and it seems more in the spirit of the assumption to compare this transition probability with $\delta_{\kappa(a_p^i(h');\zeta_t)}$. The two are not the same, because even assuming strong asymptotic empiricism, the empirical frequency of transitions from $a_p^i(h')$ to h' need not be the same asymptotically as the probability that i assesses for this transition. This unhappy possibility can arise, for example, owing to asymptotic independence: If the transition from $a_p^i(h')$ to h' requires two other players j and j' to choose (say) a and a', it can be that $\kappa(a;\zeta_t)/\kappa(h(a);\zeta_t) = 1/2$ and similarly for a', thus i assesses probability $1/4$ for the joint action a, a', yet a, a' has never occurred. The reverse is possible as well; i can asymptotically assess zero probability for this transition, even though it has occurred a positive fraction of the times it could have occurred.

Despite the nonequivalence of the two definitions, we leave the definition as is for now. In the next section, we will provide a means by which the two become "equivalent," at least within the context of the larger story we are telling.

5.3. The consequences of MME

Proposition 5.2. *Fix behavior rules for all the players. Suppose that player i's behavior rule $\hat{\pi}^i$ satisfies the MME condition. Suppose as well that for some strategy profile π_* and some $\epsilon > 0$, if $\pi_*(a) > 0$, then $\hat{\pi}_t(\zeta_t)(a) \geq \epsilon$ for all t and $\zeta \in \Lambda \subseteq \mathcal{Z}$. Then almost surely on Λ, every information set that is π_*-relevant for player i will be reached infinitely often.*

In order to prove this, and for purposes of later results, the following generalization of Proposition 5.1 is useful.

Lemma 5.1. *Fix behavior rules for all the players. If $\hat{\pi}^i$ satisfies the MME condition, then for every $h \in H^i$ and for every $a \in A(h)$,*

$$\mathbf{P}(\{\zeta : \lim_{t\to\infty} \kappa(h;\zeta_t) = \infty,\ \liminf_{t\to\infty} \kappa(h';\zeta_t)/\kappa(a_p^i(h');\zeta_t) > 0 \text{ for all } h' \in \Phi(h),\ \text{and } \lim_{t\to\infty} \kappa(a;\zeta_t) < \infty\}) = 0.$$

The proof is sufficiently similar to the the proof of Proposition 5.1 that we leave it to the reader.

Proof of Proposition 5.2. Fix player i and i's decision tree, For every terminal node $z \in Z$, let $(h_1^i(z), a_1^i(z), h_2^i(z), a_2^i(z), \ldots, h_{n(z)}^i(z), a_{n(z)}^i(z))$ be the sequence of (consecutive) information sets of i and actions taken by i that precede z in i's decision tree. (Note that $n(z)$ is a function of z, and $n(z) = 0$ if the path to z avoids all of i's information sets.)

Extend the concept of π_*-relevance to outcomes, by calling $z \in Z$ relevant if, when i takes a strategy that calls for $a_j^i(z)$ at $h_j^i(z)$, for $j = 1, \ldots, n(z)$ and i's rivals play according to π_*, then z is reached with positive probability. We will show that every relevant z will be reached infinitely often (almost surely on Λ), which suffices for the desired result.

If z is relevant, then (if i's rivals use π_*), there is positive transition probability of moving (a) to $h_1^i(z)$, (b) from $a_j^i(z)$ to $h_{j+1}^i(z)$, for $j = 1, \ldots, n(z) - 1$ and (c) from $a_{n(z)}^i(z)$ to z. Under the hypothesis of the lemma, on the event Λ there is a strictly positive lower bound γ on these different transition probabilities that applies uniformly at all dates. (The value of γ depends on the number of actions that might be resolved from one information set of i to the next, and the probability distribution on moves by nature.)

We show inductively that along the path leading to a relevant node z, every one of i's information sets is reached and every action there (and, in particular, the action that is the next step to z) is taken infinitely often. Since there is a uniform lower bound on the probability of the last step (from $a_{n(z)}^i$ to z), this gives the result. Because a formal, detailed proof takes a great deal of space, we only sketch the induction:

The probability of reaching $h_1^i(z)$ on any round is uniformly greater than γ,

so by the law of large numbers, the relative frequency that $h_1^i(z)$ is reached will have limit infimum of γ or more. Hence by Lemma 5.2, every action at $h_1^i(z)$, and in particular $a_1^i(z)$, will be taken infinitely often. Each time $a_1^i(z)$ is taken, there is probability at least γ of reaching $h_2^i(z)$, hence by Lemma 5.2 hence $a_2^i(z)$ will be taken infinitely often. If we assume inductively that $h_n^i(z)$ is reached infinitely often and that the limit infimum of each transition frequency is greater than γ in moving through the tree to $h_n^i(z)$, then Lemma 5.2 tells us that $a_n^i(z)$ is taken infinitely often. Each time $a_n^i(z)$ is taken there is probability at least γ of reaching $h_{n+1}^i(z)$, which gives the induction step. ∎

5.4. Correlated experiments and the uniform nonexperimental condition

We aim to prove results of the following form: *Given a non-Nash strategy profile* π_*, *there is an* $\epsilon > 0$ *such that for all models of beliefs and behavior in which beliefs rules are strongly asymptotically empirical, and behavior rules are asymptotically myopic with experience-time limitations on experiments and satisfy MME,*

$$\mathbf{P}(\|\check{\pi}_t(\zeta_t) - \pi_*\| < \epsilon \textit{ for all } t) = 0.$$

The argument is meant to go as follows: Suppose that the probability of the set is positive. (1) On this set of positive probability, all π_*-relevant information sets will be reached infinitely often, by virtue of Proposition 5.2. (2) The empirical record of behavior there will be close to that prescribed by π_* (by the strong law of large numbers and experience-time limitations on experimentation), and thus (3) beliefs will be close to those prescribed by π_* (by strong asymptotic empiricism). (4) Invoke Proposition 3.2 and asymptotic myopia again to get a contradiction.

This line of proof doesn't quite work, because the first step in the chain of assertions is false. Proposition 5.2 requires that $\hat{\pi}_t(\zeta_t)$ is "close" to π_*; it isn't sufficient for $\check{\pi}_t(\zeta_t)$ to be close to π_*. Consequently, it is possible that each

player's behavior rule satisfies infinite strategic-experimentation or MME and that $\check{\pi}_t(\zeta_t)$ is always close to π_*, yet there are π_*-relevant information sets that are not reached infinitely often. An example will illustrate the difficulty.

Example 5.1. Consider the game in Figure 2. (Only payoffs for players 1 and 2 are supplied.) It is not a Nash equilibrium for both players 1 and 2 to choose Across and across, respectively, with probability one; whatever they believe about the actions of player 3, one or the other would wish to deviate. However it is easy to manufacture distinct beliefs for the two of them about player 3's strategy that makes Across–across into a self-confirming equilibrium. Note that at any strategy profile that involves Across–across, player 3's information set is relevant to both players 1 and 2. So suppose, in the spirit of an assumption that each player must experiment with each of his pure strategies infinitely often, player 1 decides to try Down at dates $t = 1, 10, 100, 1000, \ldots$. Player 2, being symmetric also decides to experiment with down at dates $t = 1, 10, 100, 1000, \ldots$. Then even if the non-experimental portions of the behavior of players 1 and 2 are precisely the target strategy (Across–across), we never reach player 3's information set.

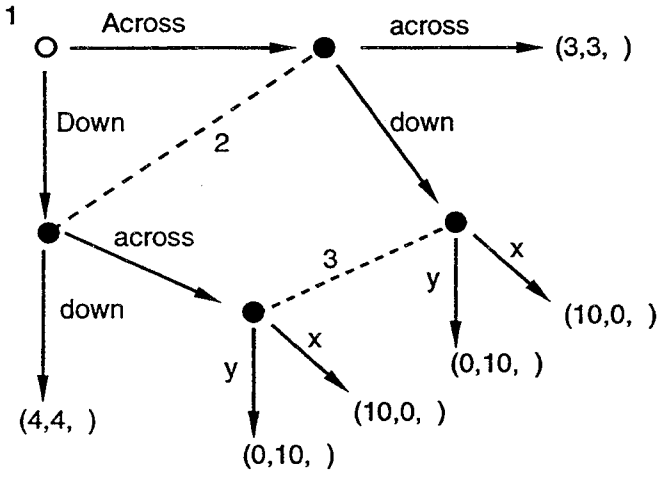

FIG. 2. Example 5.1: Correlated experiments.

The problem indicated by this example is that if experiments are sufficiently

positively correlated, they can act to frustrate one another. For an experiment by a single player to work in the sense of causing a relevant information set to occur (with positive probability), there should be positive probability (bounded away from zero) that this experiment is the only experiment being taken.

This suggests a cheap fix for this final difficulty.

Definition. *The behavior rule* $\hat{\pi}^i$ *is* ***uniformly nonexperimental*** *if, for some* $\alpha > 0$, $\check{\alpha}^i_t(\zeta_t) \geq \alpha$ *for all* t *and* ζ_t.

There is probably no ambiguity as what this means, or why it is not very attractive as a condition on behavior. It should also be fairly clear that if $\check{\pi}_t(\zeta_t)$ lies in a sufficiently close neighborhood of some target strategy π_* and if every player's behavior rule is uniformly nonexperimental, then we can employ Proposition 5.2 to get the desired result:

Proposition 5.3. *For any non-Nash strategy profile* π_*, *there exists* $\epsilon > 0$ *such that*

$$\mathbf{P}\big(\|\check{\pi}_t(\zeta_t) - \pi_*\| < \epsilon \text{ for all } t\big) = 0,$$

for any model of behavior and beliefs where the beliefs are strongly asymptotically empirical, and behavior is asymptotically myopic with maximal experience-time experimentation, uniformly nonexperimental, and satisfies the MME condition.

We will not bother with the proof. In the next section, we will see how to dispense with uniform nonexperimentation in a manner that brings a number of other benefits; and then in Section 7 we will get a result in the general nature of Proposition 5.3. The proof of this result (Proposition 7.1) will give enough details that the steps of the proof should be clear.

6. Statistical tests

Our two fundamental behavioral assumptions, asymptotic empiricism and asymptotic myopia, are most sensible if each player believes that his opponents'

behavior converges to the repeated play of some fixed strategy profile. More formally, each player should believe that his opponents' play is asymptotically exchangeable and independent (of the play of others).

In our formalism of asymptotic empiricism, we implicitly assume that players maintain this basic hypothesis regardless of the strength of evidence to the contrary. It seems to us that players would not be so doctrinaire in their beliefs. They might, at the outset, entertain as a working hypothesis that their rivals' behavior is asymptotically exchangeable and independent, but this working hypothesis, together with the strictures of asymptotic empiricism and asymptotic myopia, would be discarded when and if there is sufficient evidence against it.

To model this possibility, we suppose that players run statistical tests of the working hypothesis, rejecting the hypothesis given sufficient evidence against it. We modify the definitions of stability (i.e., of unstable and weakly stable profiles), to include as a condition of stability that observed play passes the players' tests of exchangeability and independence. This allows us to use less restrictive versions of asymptotic empiricism and asymptotic myopia, and to dispense with uniform nonexperimentation.

6.1. Statistical test sequences

To this point, in any learning model player i has been fully described by his beliefs rules $\hat{\gamma}^i$ and behavior rule $\hat{\pi}^i$. To incorporate statistical tests into our formulation, to the specification of player i we add a sequence $\{\Lambda_t^i; t = 1, 2, \ldots\}$ where: (a) each Λ_t^i is a ζ_t-measurable subset of $\mathcal{Z}$, i.e., whether $\zeta \in \mathcal{Z}$ is or is not an element of Λ_t^i is determined by the history up to time t in ζ; and (b) $\Lambda_t^i \subseteq \Lambda_{t-1}^i$ for all $t > 1$. The interpretation is that if $\zeta \in \Lambda_t^i$, then along the history ζ, player i at date t has not yet rejected the hypothesis of asymptotically empirical and independent behavior by his rivals; if $\zeta \notin \Lambda_t^i$, then this hypothesis has been rejected by i at or before t. Rejection must be based on the data available to i, hence (a). And the basic model, once rejected, cannot

be "unrejected," hence (b).[14] We call a specification of $\Lambda^i = \{\Lambda^i_t; t = 1, 2, \ldots\}$ a *statistical test sequence* for i.

Definitions. *(a) Player i's belief rule $\hat{\gamma}^i$ is* ***strongly asymptotically empirical (relative to the statistical test sequence Λ^i)*** *if for every $\epsilon > 0$, every infinite history ζ such that $\zeta \in \cap_{t=1}^{\infty} \Lambda^i_t$, every information set $h^j \in H_{\text{i.o.}}(\zeta) \cap H^j$ for $j \neq i$, and every $a \in A(h^j)$,*

$$\lim_{t\to\infty} \hat{\gamma}^i_t(\zeta_t)\left(\left\{\pi^{-i} \;:\; \left|\pi^j(a) - \frac{\kappa(a;\zeta_t)}{\kappa(h^j;\zeta_t)}\right| < \epsilon\right\}\right) = 1.$$

(b) Fix player i and i's beliefs rule $\hat{\gamma}^i$. The behavior rule $\hat{\pi}^i$ for i is ***asymptotically myopic with experience-time limitations on experimentation (with respect to the statistical test sequence Λ^i)*** *if it satisfies the definition of asymptotic myopia with experience-time limitations given previously, but with every appearance (in the definition) of ζ_t weakened to be, for $\zeta_t \in \Lambda^i_t$.*[15]

(c) The behavior rule $\hat{\pi}^i$ of player i satisfies ***MME (with respect to the statistical test sequence Λ^i)*** *if it satisfies the definition of MME given previously, but where the defining conditions need hold only for $\zeta_t \in \Lambda^i_t$.*

Perhaps most significantly, the definitions of unstability and local stability are changed:

Definitions. *(a) The strategy profile π_* is* ***unstable*** *for a given class of learning models if there exists $\epsilon > 0$ such that, for all learning models from this class,*

$$\mathbf{P}\left(\|\breve{\pi}_t(\zeta_t) - \pi_*\| < \epsilon \text{ and } \zeta_t \in \Lambda^i_t \text{ for all } t \text{ and } i\right) = 0.$$

[14] The latter is for expositional simiplicity and can be modified.

[15] The condition $\zeta_t \in \Lambda^i_t$ is a bit abusive of notation, but the meaning should be clear.

*(b) The outcome ρ_* is **unstable** for a given class of learning models if there exists $\epsilon > 0$ such that, for all learning models from this class,*

$$\mathbf{P}\big(\|\rho(\check{\pi}_t(\zeta_t)) - \rho_*\| < \epsilon \text{ and } \zeta_t \in \Lambda_t^i \text{ for all } t \text{ and } i\big) = 0.$$

*(c) The strategy profile π_* is **locally stable** for a given class of learning models if for some learning model from the given class,*

$$\mathbf{P}\big(\lim_{t\to\infty} \check{\pi}_t(\zeta_t) = \pi_* \text{ and } \zeta_t \in \Lambda_t^i \text{ for all } t \text{ and } i\big) > 0.$$

That is, if π_* is unstable, then for all learning models (in the specified class), at some point either behavior diverges from π_* or some player rejects the working hypothesis, almost surely. For local stability, there must be positive probability of (1) convergence to π_* and (2) no rejection of the working hypothesis.

Adding statistical test sequences to the basic story according to the definitions above has two obvious effects:

(1) Insofar as the conditions for asymptotic empiricism, asymptotic myopia, and minimum experimentation are required only for histories that pass the statistical tests imposed, those conditions become weaker and therefore more palatable.

(2) Conversely, the addition of statistical test sequences to the definition of unstable profiles and outcomes increases the set of histories where behavior is unstable, giving a weaker definition. Hence the results that non-Nash profiles and/or outcomes are unstable become weaker when statistical tests sequences are invoked.

In view of these two effects, we seek classes of statistical test sequences that make our assumptions on behavior more palatable and remove the need for unpalatable assumptions — e.g., the uniform nonexperimentation — while at the same time not overly or implausibly restricting the set of stable profiles/outcomes.

Note that the general formulation described above does not address how players behave when and if some players' statistical test fails. When a profile is unstable, and its unstability depends on the failure of a statistical test, we do not know that behavior cannot remain in some small neighborhood of the target strategy. Rather, all we can conclude is that if behavior doesn't eventually move (a finite distance) away from the target, then one or more player's observations will look "odd" for a sequence generated by asymptotically exchangeable and independent play.

6.2. Arc-frequency statistical test sequences

We now propose a particular type of statistical test sequence, based on a fairly simple and direct test of asymptotic exchangeability and independence, which helps give us the results we want.

To prepare for the definition, we invent more notation: For $v \in V \setminus W$, let $x_p(v)$ be the node (in X) that immediately precedes v, and let $a_p(v)$ be the action that leads from $x_p(v)$ to v. Also, let $i_p(v)$ be the player who moves at node $x_p(v)$; i.e., $i_p = i(h(x_p(v)))$.

Definition. *A statistical test sequence $\{\Lambda_t^i\}$ for player i contains an* ***arc-frequency test*** *if, for some sequence of nonnegative numbers $\{\eta_n\}$ with $\liminf_n \eta_n > 0$ and some $\gamma^* \in (0,1)$,*

$$\zeta_t \in \Lambda_t^i \quad \textit{implies that for all } v \in V \setminus W \textit{ such that } i_p(v) \neq i,$$
$$\hat{\gamma}_t^i(\zeta_t)\left(\left\{\pi^{-i} \in \Pi^{-i} \;:\; \kappa(v;\zeta_t) \geq \kappa(x_p(v);\zeta_t)[\eta_{\kappa(x_p(v);\zeta_t)}\pi^{i_p(v)}(a_p(v))]\right\}\right) \geq 1-\gamma^*. \tag{6.1}$$

To make sense of this definition, consider the contrapositive of (6.1): If

$$\hat{\gamma}_t^i(\zeta_t)\left(\left\{\pi^{-i} \in \Pi^{-i} \;:\; \kappa(v;\zeta_t) < \kappa(x_p(v);\zeta_t)\pi^{i_p(v)}(a_p(v))\eta_{\kappa(x_p(v);\zeta_t)}\right\}\right) > \gamma^*$$

for some $v \in V$ with $i_p(v) \neq i$, then the statistical test fails. (Since we assume that the Λ_t^i are nested, the test fails at t' if this is true for any $t \leq t'$.) In

the test, the number of times the arc from $x_p(v)$ to v has been traversed is compared with the number of times this transition should have been made, if $i_p(v)$ takes action $a_p(v)$ with the probability $\pi^{i_p(v)}(a_p(v))$, for each π^{-i}. Thinking η_n as being fairly small for small indices and approaching some value strictly less than one as $n \to \infty$, the test fails only when this arc has been traversed "too infrequently" if i's rivals play π^{-i} repeatedly (and independently) from round to round. When i assesses probability γ^* or more for the set of π^{-i} that give too few traversals of any arc, then i's beliefs about π^{-i} are too-highly concentrated on values of π^{-i} which, the data indicate, do not support the basic hypothesis of exchangeable and independent play by i's rivals.

Several remarks are in order:

(1) One might wonder why asymptotic empiricism doesn't make this condition moot: How could player i's beliefs after many observations continue to assign substantial probability to strategy profiles that give "too few" observations of certain arcs? The answer, and the reason this test is needed, is that asymptotic empiricism implicitly assumes asymptotic exchangeability and independence. The test asks whether the data are consistent with this implicit maintained hypothesis.

(2) The test compares the actual number of times each noninitial node v was reached with an estimate that is computed assuming that player $i_p(v)$ used the same strategy in each previous period. But the hypothesis being tested is only that players asymptotically converge to repeated play of some fixed strategy. In this regard, note that it is acceptable for there to be some very large N such that the "scaling" constant $\eta_n = 0$ for $n < N$, and $\eta_n \leq (n - N)/(2n)$, say, for $n > N$, which leaves a lot of slack.

(3) In these test sequences, player i computes the probability he would assess (under the maintained hypothesis) for the set of π^{-i} that (under this hypothesis) look odd given the data. When this probability is sufficiently high, he rejects the

maintained hypothesis. An alternative test would be to compute a single estimate for the (approximate) number of times that each arc in the tree should have been traversed. Specifically, player i could first compute, for $a \in A^j$, $j \neq i$,

$$r(a; \hat{\gamma}_t^i(\zeta_t)) = \int_{\Pi^{-i}} \pi^j(a) \hat{\gamma}_t^i(\zeta_t)(\mathrm{d}\pi^{-i})$$

and then reject the maintained hypothesis if, for any v such that $i_p(v) \neq i$,

$$\kappa(v; \zeta_t) < \kappa(x_p(v); \zeta_t) r(a_p(v); \hat{\gamma}_t^i(\zeta_t)) \eta'_{\kappa(x_p(v); \zeta_t)},$$

for an appropriate sequence $\{\eta'_n\}$ with strictly positive limit infimum. Without going into detail, we remark that such tests are asymptotically almost equivalent to the sort of test we have posed, whenever i's beliefs rule is strongly asymtotically empirical. This is so because the tests only "matter" as $\kappa(x_p(v); \zeta_t)$ goes to infinity, in which case strong asymptotic empiricism implies that $\hat{\gamma}_t^i(\zeta_t)$ at $h(x_p(v))$ will be nearly a point distribution. (To make this precise, one needs to add uniformity in $\kappa(h; \zeta_t)$ to the convergence part of strong asymptotic empiricism.)

To illustrate the impact of arc-frequency statistical test sequences, we consider two examples. The first is Example 5.1. Recall that in this example, players 1 and 2 choose Across and across (respectively) with probability approaching one, although each experiments infinitely often with Down/down. Because of perfect correlation in the experiments, neither Across–down or Down–across are every observed. This history will certainly fail any arc-frequency statistical test sequence applied by player 1: player 1 tests whether, in those (infinitely many) instances where he chooses Down, is the asymptotic fraction of Down–across anything like the frequency it "ought to be," namely 1. Of course, it is not; we could only pass an arc-frequency statistical test sequence for player 1 if, asymptotically, Down–across happens a strictly positive fraction of the time that Down occurs. Thus we see a way in which arc-frequency statistical test sequences test

for correlation in the play of the player running the test and the *subsequent* play of his rival. [16]

For the second example, imagine player 3 who observes the behavior of players 1 and 2 in an extensive-form game where 1 chooses between Left and Right and 2 between left and right, simultaneously and independently. Suppose that player 1 chooses Left half the time, and player 2 chooses left half the time. Then in order to pass any arc-frequency statistical test sequence employed by player 3, Left–left must occur a positive fraction of the time. And if the arc-frequency statistical test that is employed has $\lim_n \eta_n = 1$, then Left–left must occur precisely 1/4 of the time (in the limit), or the test will fail: If Left–left occurs more than 1/4 of the time, the test will fail for some other of the three possible combinations. [17]

6.3. Arc-frequency tests and asymptotic myopia

The previous example indicates how arc-frequency statistical test sequences make more palatable the assumption of strong asymptotic empiricism. The same is true about asymptotic myopia, at least with reference to the "troublesome example" of Fudenberg and Kreps (1994, Section 4). Recall that the problem there arose because asymptotic independence led a player to believe that, if he chose a given action, an information set down the tree would be reached with positive probability. Knowledge of what would happen at that information set mattered to his decision whether to take the action, which might lead him to experiment with the action. But (along the particular sample path described in the example)

---

[16] The emphasized subsequent will be explained below.

[17] To prevent too-rapid rejection of the basic hypothesis when it is true, one doesn't want $\eta_n = 1$, but rather $\eta_n = 1 - 1/(o(n))$ for some $o(n)$ function. For example, the law of the iterated logarithm would suggest something like $o(n) = O(n^{1/2} \ln\ln(n))$. Note that if we define $\eta_n = \max\{0, 1 - K/[n^{1/2} \ln\ln(n)]$, then as $K \to \infty$ there is vanishingly small probability of rejecting the hypothesis if it is true, while (for fixed K) asymptotic independence is no longer an asymptotically troublesome aspect of asymptotic empiricism.

every time he tried this action, the information set in question was not reached. Asymptotic myopia required him to experiment with this action less and less frequently, even though (until the information set is reached) the value of the information expected to be obtained remains sufficient (in a discounted present-value-of-expected-payoffs criterion) to make this action optimal. Compare this with precise value-of-information calculations in multi-armed bandit problems with independent arms and nondoctrinaire priors, where the information value of trying a particular arm falls to zero uniformly in the number of times that the arm is pulled. That result fails to extend to this example because, although the player keeps experimenting, the information that (by asymptotic independence) is expected to be generated never is generated. *If* we were ensured that information sets that ought to be reached with positive probability (given an experiment by player i) are reached with a number of times that goes to infinity in the number of times the information set "ought" to be reached, then this example would be mooted. Since no such guarantee is possible, an alternative tack would be to suppose that whenever the information set is not being reached infinitely often (so that the value of information remains high), the limitations on experimentation that come with asymptotic myopia are suspended.

It is this second, alternative tack that is engaged with an arc-frequency statistical test. If i believes (asymptotically, or even along some subsequence of dates) that an information set can be reached if he tries to get there, and he *does* try to get there, yet he is repeatedly frustrated in these attempts, then the test will eventually fail, and the strictures of asymptotic myopia dissolve.

6.4. Noninvariance to extensive form and Own-action-independence statistical test sequences

The power of arc-frequency statistical tests can depend on "strategically irrelevant" features in the extensive form of the game, such as the interchange of simultaneously moves. We can see this in Example 5.1, for the history described

in the example. As noted already, along this history player 1's statistical test sequence will reject, as long as 1's statistical test sequence contains an arc-frequency test. But this is not true for player 2, even though the game depicted is symmetric (in terms of the strategic form) in players 1 and 2. Player 1 is able to compute the "conditional frequency" of down, conditional on he himself choosing Down, since (in the game tree) 2's choice follows his own. But, at least in terms of the formal definition of an arc-frequency test, 2 does not compute the conditional frequency of Down, conditional on down.

We believe it is possible to reformulate arc-frequency statistical tests in a way that is appropriately invariant to the extensive form representation of the game. But it is easier expositionally to complement arc-frequency statistical tests as formulated with a different type of statistical test, again based on a given extensive form of the game, which (together with arc-frequency statistical tests) is adequate for our purposes. We will follow the path of expositional ease.

Definition. *A statistical test sequence* $\{\Lambda_t^i\}$ *for player* i *contains an own-action-independence test if, for some sequence of nonnegative numbers* $\{\eta_n\}$ *with* $\liminf_n \eta_n > 0$*: For every* t, $\zeta_t \in \Lambda_t^i$ *implies that for all* x *and* a *such that* $h(x) \in H^i$ *and* $a \in A(h)$,

$$\frac{\kappa((x,a);\zeta_t)}{\kappa(a;\zeta_t)} \geq \frac{\kappa(x;\zeta_t)}{\kappa(h(x);\zeta_t)} \eta_{\kappa(x;\zeta)}.$$

The idea is relatively straightforward. Suppose $\eta_n = 0$ for, say, $n \leq$ 100,000, and $\eta_n = .5$ thereafter. Then when and if node x has been visited 100,000 times (so information set $h(x)$ has been visited at least this many times), the ratio of the number times a was taken at x to the number of times a was taken overall should be at least one-half the ratio of the number of times x has been visited to the number of times $h(x)$ was visited. If the former ratio remains very much less than the latter, then somehow i's rivals' actions preceding $h(x)$ are correlated with i's presumably independent choice of action at $h(x)$.

6.5. *Other statistical test sequences*

We will work hereafter with models of beliefs and behavior in which all the players employ statistical test sequences that incorporate both arc-frequency and own-action-independence tests. This, it turns out, is adequate to the results we are aiming for. But we do not mean to imply that only arc-frequency/own-action-independence statistical test sequences are sensible. Statistical test sequences could be used which (also) test for cycles in the behavior of a single player, or for a general lack of convergence in behavior. [18] (For example, one tests for large N whether behavior over the first $N/2$ periods is similar to behavior over the rest of history.) Of course, as one adds more and more tests, one can feel better and better about asymptotic empiricism and myopia. But equally of course, since more tests makes failure easier, and since unstability of a profile or outcome is triggered if ever a test is failed, this weakens any results about the unstability of a given profile or outcome.

7. *Stability and Nash equilibrium profiles*

Proposition 7.1. *If π_* is not a Nash equilibrium profile, then π_* is unstable for the class of learning models in which: Beliefs rules are strongly asymptotically empirical; behavior rules satisfy asymptotic myopia with experience-time limitations on experimentation and MME; and each player employs a statistical test sequence which incorporates both arc-frequency and own-action-independence tests.*

Proof. Suppose that π_* is not a Nash equilibrium strategy profile. Let

$$\epsilon = \min\left\{\epsilon', \min_{\{a \in A: \pi_*(a) > 0\}} \{\pi_*(a)\}\right\}/2,$$

where ϵ' is the ϵ' produced for π_* according to Proposition 3.2. Let

$$\Lambda = \left\{\zeta \in \mathcal{Z} : \|\check{\pi}_t(\zeta_t) - \pi_*\| \leq \epsilon \text{ and } \zeta_t \in \Lambda_t^i \text{ for all } t \text{ and } i\right\},$$

[18] For different approaches to cyclic behavior, see Aoyagi (1992) and Sonsino (1994).

and suppose that for some learning model from the class described, $\mathbf{P}(\Lambda) > 0$.

Because all the behavior rules satisfy MME, there is zero probability for each of the events described in the statement of Lemma 5.1. For all h and $a \in A(h)$,

$$\mathbf{P}\left(\left\{\zeta : \lim_t \kappa(h;\zeta_t) = \infty, \limsup_t \left|\frac{\kappa(a;\zeta_t)}{\kappa(h;\zeta_t)} - \pi_*(a)\right| > \epsilon\right\} \cap \Lambda\right) = 0,$$

by the strong law of large numbers and the experience-time experimentation condition. There is, of course, probability one that every $w \in W$ is hit infinitely often.

Let Λ^0 be the set Λ after discarding all these zero-probability events; i.e., Λ^0 is a set of positive measure on which every $w \in W$ is hit infinitely often, players' empirical frequencies of actions asymptotically are within ϵ of π_* at information sets visited infinitely often, players take every action infinitely often at any information set h that is visited infinitely often and that has non-zero lim infs for all the transitition probabilities needed to reach h, all the statistical tests are passed, and $\breve{\pi}_t$ is always within ϵ of π_*.

We claim that for every $x \in X$ that is π_* relevant to some player i and for every $\zeta \in \Lambda^0$: (a) x is reached infinitely often; and (b) if $h(x) \in H^i$, then

$$\liminf_{t\to\infty} \frac{\kappa(x;\zeta_t)}{\kappa(h(x);\zeta_t)} > 0 \text{ and } \liminf_{t\to\infty} \frac{\kappa(h';\zeta_t)}{\kappa(a(h');\zeta_t)} > 0 \text{ for all } h' \in \Phi(h(x)).$$

We prove this by induction on the length of the path leading to x. For all initial nodes, the result is true by the construction of Λ^0. So suppose the result is true for all π_*-relevant nodes that are n steps or less from an initial node, and select some π_*-relevant node x that is $n+1$ steps from an initial node. This node is relevant to some player, whom we denote by i. Until further notice, let x' denote the immediate predecessor of x, and let a denote the action that leads from x' to x.

Of course, x' is π_*-relevant to i and is n steps from an initial node, so we know that x' (and all of its predecessors) must have be reached infinitely

often along each history $\zeta \in \Lambda^0$. Thus for each predecessor node of x which belongs to an information set that doesn't belong to i, i's beliefs concerning the actions taken at the corresponding information set must be converging to a point mass at the empirical frequency of actions, which (for large enough t) must be within ϵ of π_* at that information set. Fixing a history ζ, assume that T is sufficiently large so that for all $t \geq T$, $\hat{\gamma}_t^i(\zeta)$ ascribes probability greater than $1 - \gamma^*/2$ to strategies that are no more than $\epsilon/4$ from the empirical frequency (at each of these information sets), and such that the empirical frequencies are no more than $5\epsilon/4$ away from π_*. Because ζ passes all the statistical tests applied by i, and in particular the arc-frequency tests, this implies that for large enough t, the frequency of transitions along the path to x' that are not in i's control must be strictly positive and uniformly bounded away from zero. (The precise frequency of these transitions is determined by the size of $\min \pi_*(a)$ and the lim inf of the $\{\eta_n\}$ sequence in the arc-frequency portion of the statistical test.)

If $h(x') \in H^{-i}$, the argument just given shows that the frequency of transitions out of x' that are a must be a strictly positive fraction of all transitions out of x'. Since x' is visited infinitely often, so must be x.

If $h(x') \in H^i$, $\liminf_{t\to\infty} \kappa(x';\zeta_t)/\kappa(h(x');\zeta_t) > 0$ by the induction hypothesis. Moreover, because $h(x')$ is visited infinitely often and $\kappa(h';\zeta_t)/\kappa(a(h');\zeta_t)$ has strictly positive lim inf for all $h' \in \Phi(h(x'))$, action a is necessarily taken infinitely often by i. Because ζ passes the own-action-independence statistical tests, $\kappa((x',a);\zeta_t)/\kappa(a;\zeta_t)$ must have strictly positive lim inf, and thus $\kappa((x',a);\zeta_t)$ must go to infinity. Of course, $\kappa((x',a);\zeta_t)$ is just $\kappa(x;\zeta_t)$.

Thus we have shown (a) of the induction hypothesis. It remains to show part (b). Accordingly, suppose that $h(x) \in H^i$. Let h' denote information set belonging to player i that immediately precedes $h(x)$ in i's decision tree, let x' now denote the (unique) element of h' that precedes x, and let a' be the (unique) action that constitutes the first step from x' to x. (If $h(x)$ is an initial

information set for i, a small modification of the argument is needed.) Since all the transitions from (x', a') to x are controlled by players other than i and x is π^*-relevant to i, the arc-frequency statistical tests ensure us that

$$\liminf_{t\to\infty} \frac{\kappa(x;\zeta_t)}{\kappa((x',a');\zeta_t)} > 0.$$

(Recall that we have already proven that x, and thus each step along the way, is visited infinitely often.) By perfect recall, $\kappa(h;\zeta_t) \leq \kappa(a';\zeta_t)$, and thus

$$\frac{\kappa((x',a');\zeta_t)}{\kappa(h;\zeta_t)} \geq \frac{\kappa((x',a');\zeta_t)}{\kappa(a';\zeta_t)}.$$

By own-action independence,

$$\frac{\kappa((x',a');\zeta_t)}{\kappa(a';\zeta_t)} \geq \frac{\kappa(x';\zeta_t)}{\kappa(h';\zeta_t)}^{\nu_{\kappa(x';\zeta_t)}},$$

which has strictly positive lim inf by the induction hypothesis. Combining these inequalities, we see that

$$\liminf_{t\to\infty} \frac{\kappa(x;\zeta_t)}{\kappa((x',a');\zeta_t)} \cdot \frac{\kappa((x',a');\zeta_t)}{\kappa(h;\zeta_t)} = \frac{\kappa(x;\zeta_t)}{\kappa(h;\zeta_t)} > 0.$$

As for the second half of (b), we only need to worry about

$$\liminf_{t\to\infty} \frac{\kappa(h;\zeta_t)}{\kappa(a';\zeta_t)} > 0;$$

for all other $h'' \in \Phi(h(x))$, the induction hypothesis applies. And for this final asymptotic inequality, write

$$\frac{\kappa(h;\zeta_t)}{\kappa(a';\zeta_t)} \geq \frac{\kappa(x;\zeta_t)}{\kappa(a';\zeta_t)} = \frac{\kappa(x;\zeta_t)}{\kappa((x',a');\zeta_t)} \cdot \frac{\kappa((x',a');\zeta_t)}{\kappa(a';\zeta_t)}.$$

We know that each of the ratios on the right-hand side has strictly positive lim inf, so their product does as well.

The rest is simple. For every $\zeta \in \Lambda^0$, every node x that is relevant to some player is hit infinitely often. Hence beliefs at all those nodes, and hence at the

information sets they contain, converge to the empirical frequencies, which must be close enough to π_* so that (at a sufficiently late date) some player is forced by asymptotic myopia to choose a nonexperimental portion of his strategy that is more than ϵ away from π_*, contradicting the definition of Λ^0. ∎

Proposition 7.2. *If π_* is a Nash equilibrium profile, then π_* is weakly stable for the class of learning models in which: Beliefs rules are asymptotically empirical; behavior rules satisfy asymptotic myopia with experience-time limits on experimentation and MME; and each player employs a statistical test sequence that includes both arc-frequency and own-action independence statistical tests.*

As was the case with similar results in Fudenberg and Kreps (1993, 1994), the construction is very artificial and not very informative. Hence we will be content to sketch one way of showing this. Let I be the number of players and let M be the maximal number of times any player can move during a single round of the game. We construct strategies in which, at each date t, at most one player will be experimenting: Player 1 will experiment at dates $kI+1$ (only), player 2 at dates $kI+2$, and so on. On dates when players do not experiment, they play according to the given Nash equilibrium π_*. On date $kI+i$, player i randomizes independently at each information set whether to experiment there or not. With probability $(1/t)^{1/M}$, i experiments at (any given) information set h, where an experiment consists of trying each available action with equal probability. With complementary probability, i plays π_* at h.

Suppose players use the behavior rules just described. By a relatively straightforward application of the Borel-Cantelli lemma, we can show that every π_*-relevant information set will be reached infinitely often. By the strong law of large numbers, empirical frequencies of play at every information set that is reached infinitely often will converge to the prescriptions of π_*. By a somewhat more complex argument, we show that: For any node x belonging to player

i that is π_* relevant to i, $\kappa(x;\zeta_t)/\kappa(h(x);\zeta_t)$ is either identically zero or it has strictly positive limit, and if the limit is strictly positive, then for every action $a \in A(h)$, $\kappa((x,a);\zeta_t)/\kappa(a;\zeta_t)$ has the same limit.

Suppose player i had beliefs that put probability one on rivals playing π_*^{-i}. Then the behavior rule described above, put in the role of $\check{\pi}^i$, would satisfy asymptotic myopia (since experiments, when they are taken, are taken with vanishingly small probability). The proof then proceeds as follows: Use each of the a.s. limits described in the previous paragraph to find a set of positive probability on which all the limits are achieved in uniform fashion. Use the estimates in those uniform convergences to define strong asymptotic empiricism, asymptotic myopia, and the own-action-independence statistical test so that, on this set of positive probability, players are not required by MME to experiment more than is prescribed by $\check{\pi}^i$, nor abandon the belief that their rivals are surely playing π_*^{-i}, nor reject on the basis of own-action-independence. In this regard, note that non-relevant information sets are never reached, so MME can be specified so that no experiments at these information sets are ever required. In similar fashion, the triggers of the arc-frequency statistical tests can be specified so that the tests are passed.[19] On the set where any of the uniform limits fails to hold, redefine beliefs and behavior in any fashion consistent with the requirements of the proposition. Since there is positive probability that no resetting is necessary, there is no change in the probability of and probabilities within the set where behavior (and beliefs) are unchanged, which gives the result, because the behaviors described satisfy $\check{\pi}^i \to \pi_*^i$.

8. Stability and Nash equilibrium outcomes

In this section, we discuss the following result:

[19] We put arc-frequency statistical tests last because they cannot be conducted until beliefs are specified.

Proposition 8.1. *If ρ_* is not a Nash equilibrium outcome, then ρ_* is unstable for the class of learning models in which: Beliefs rules are strongly asymptotically empirical; behavior rules satisfy asymptotic myopia with calendar-time limitations on experimentation and MME; and each player employs a statistical test sequence which incorporates both arc-frequency and own-action-independence tests.*

Comparing with Proposition 7.1, the difference is that experimentation at any information set is limited by calendar time instead of experience time at that information set. This means that behavior at infrequently-reached information sets can be capricious, a point to which we return after giving the proof of the proposition.

To prove this proposition, we begin with some notation and two lemmas.

For any outcome ρ_*, let $\bar{H}(\rho_*)$ denote information sets that are reached with positive probability under ρ_*. For $h \in \bar{H}(\rho_*)$ and $a \in A(h)$, let $\pi_*(a) = \rho_*(Z(a))/\rho_*(Z(h))$. That is, π_* is a partial strategy profile defined (only) for information sets $h \in \bar{H}(\rho_*)$.

Lemma 8.1. *Fix any outcome ρ_* that corresponds to some strategy profile for the stage game, and let π_* be the partial strategy profile defined from ρ_* as above. Then:*

(a) $\rho(\pi) = \rho_*$ if and only if $\pi(a) = \pi_*(a)$ for all $a \in A(\bar{H}(\rho_*))$;

(b) for every $\epsilon > 0$, there is a $\delta > 0$ such that if $\|\pi(h) - \pi_*(h)\| \leq \delta$ for every $h \in \bar{H}(\rho_*)$, then $\|\rho(\pi) - \rho_*\| \leq \epsilon$; and

(c) for every sufficiently small $\epsilon > 0$, there is a $\delta > 0$ such that if $|\pi(a) - \pi_*(a)| \geq \delta$ for any $a \in A(\bar{H}(\rho_*))$, then $\|\rho(\pi) - \rho_*\| \geq \epsilon$.

This lemma essentially involves careful bookeeping, and the details are omitted.

Lemma 8.2. *If ρ_* is not a Nash equilibrium outcome, then for some $\epsilon' > 0$: For every strategy profile π_* such that $\|\rho(\pi_*) - \rho_*\| < \epsilon'$ and for all beliefs $\{\gamma^i\}$ such that*

$$\gamma^i(\{\pi^{-i} : \|\pi(h) - \pi_*(h)\| \leq \epsilon' \text{ for all } h \in H^{-i} \text{ that are } \pi_* - \text{relevant to } i\}) > 1 - \epsilon',$$

there is some player i *such that*

$$\min_{\tilde{\pi}^i, \hat{\pi}^{-i}} \{\|\rho(\tilde{\pi}^i, \hat{\pi}^{-i}) - \rho_*\| \ : \ u^i(\tilde{\pi}^i, \gamma^i) + \epsilon' \geq \max_{s^i \in S^i} u^i(s^i, \gamma^i)\} > \epsilon'.$$

Proof. Suppose that ρ_* is a profile for which no such ϵ' exists. Then for each n we can find a strategy profile π_n, a beliefs profile γ_n, a second strategy profile $\tilde{\pi}_n$, and, for each i, a partial strategy profile $\hat{\pi}_n^{-i}(i)$, such that for all i:

$$\|\rho(\pi_n) - \rho_*\| \leq \frac{1}{n}, \tag{8.1a}$$

$$\gamma^i(\{\pi^{-i} : \|\pi(h) - \pi_n(h)\| \leq \frac{1}{n} \text{ for all } h \in H^{-i} \text{ that are } \pi_n - \text{relevant to } i\}) > \frac{n-1}{n}, \tag{8.1b},$$

$$u^i(\tilde{\pi}_n^i, \gamma_n^i) + \frac{1}{n} \geq \max_{s_i \in S_i} u^i(s^i, \gamma_n^i), \text{ and} \tag{8.1c}$$

$$\|\rho(\tilde{\pi}_n^i, \hat{\pi}_n^{-i}(i)) - \rho_*\| \leq \frac{1}{n}. \tag{8.1d}$$

Looking along a subsequence if necessary, we can assume that $\lim_n \pi_n$ exists — write π_* for this limit — and for each i $\lim_n \gamma_n^i$ exists — write γ_*^i for this limit. Then continuity of the $\rho(\cdot)$ function and (8.1a) imply that

$$\rho(\pi_*) = \rho_*. \tag{8.2a}.$$

The set of π-relevant information sets is lower-semi-continuous in π, so passing to the limit in (8.1b) tells us that

$$\gamma_*^i(\{\pi^{-i} : \pi(h) = \pi_*(h) \text{ for all } h \in H^{-i} \text{ that are } \pi_* \text{-relevant to } i\} = 1. \tag{8.2b}.$$

Inequality (8.1d) ensures us that $\lim \tilde{\pi}^i(h) = \pi_*^i(h)$ for all information sets $h \in H^i$ that are along the path of play according to the outcome ρ_*, and since the *ex ante* expected utility of a player is determined entirely by his actions along the

path of play and the actions of others at the information sets that are relevant to him, this together with (8.1c) and (8.2b) imply that

$$u^i(\pi_*^i, \pi_*^{-i}) = \max_{s^i \in S^i} u^i(s^i, \pi_*^{-i}).$$

Thus π_* is a Nash equilibrium and (hence) ρ_* is a Nash equilibrium profile. ∎

Proof of Proposition 8.1. Suppose that ρ_* is not a Nash equilibrium outcome. We may assume that ρ_* is the outcome corresponding to some strategy profile π_*; otherwise $\min_{\pi \in \Pi} \|\rho(\pi) - \rho_*\| > 0$, and we can take ϵ to be half of this minimum. Let ϵ' be the corresponding value produced in Lemma 8.2. Choose δ and ϵ sufficiently small so that:

(a) $\epsilon < \epsilon'/2$;

(b) per Lemma 8.1(b), if $\|\pi(h) - \pi_*(h)\| \leq \delta$ for every $h \in \bar{H}(\rho_*)$, then $\|\rho(\pi) - \rho_*\| \leq \epsilon'/2$; and

(c) per Lemma 8.1(c), if $\|\rho(\pi) - \rho_*\| \leq \epsilon$, then $|\pi(a) - \pi_*(a)| \leq \delta/2$ for every $a \in A(\bar{H}(\rho_*))$.

Suppose by way of contradiction that for some learning model satisfying the various conditions,

$$\mathbf{P}(\|\rho(\breve{\pi}_t(\zeta_t)) - \rho_*\| < \epsilon \text{ and } \zeta_t \in \Lambda_t^i \text{ for all } t \text{ and } i) > 0.$$

Let Λ be this set of positive measure. We know from (c) that $|\breve{\pi}_t(\zeta_t)(a) - \pi_*(a)| \leq \delta/2$ for all $\zeta \in \Lambda$. A straightforward induction argument will show that, almost surely on Λ, every information set in $\bar{H}(\rho_*)$ will be hit a nonvanishing fraction of the time and (thus), with calendar-time limitations on experimentation, the empirical frequencies of actions taken at information sets $h \in \bar{H}(\rho_*)$ will have limits sup and inf within $\delta/2$ of π_*. Let Λ^0 be the subset of Λ of histories where

$h \in \bar{H}(\rho_*)$ are hit infinitely often and the limits sup and inf of actions taken at those information sets are within $\delta/2$ of π_*, intersected with the complement of the event described in the display of Lemma 5.1. Then the probability of Λ^0 equals the probability of Λ.

Fix any $\zeta \in \Lambda^0$. Looking along a subsequence if necessary, $\kappa(a;\zeta_t)/\kappa(h;\zeta_t)$ converges to some limit $\pi_\zeta(a)$, for each h and $a \in A(h)$. Note well that π_ζ is defined history by history, so for information sets $h \notin H_{\text{i.o.}}(\zeta)$, this limit is the ratio of two finite numbers. We claim that for every x that is π_ζ-relevant to some player i, (a) x is reached infinitely often; and (b) if $h(x) \in H^i$, then *along the subsequence,*

$$\liminf_{t\to\infty} \frac{\kappa(x;\zeta_t)}{\kappa(h(x);\zeta_t)} > 0 \text{ and } \liminf_{t\to\infty} \frac{\kappa(h';\zeta_t)}{\kappa(a(h');\zeta_t)} > 0 \text{ for all } h' \in \Phi(h(x)).$$

The same proof by induction that worked in the proof of Proposition 7.1 is enlisted here. The key is that asymptotic empiricism tells us that for any node x that is reached infinitely often, all players' beliefs on $H(x)$ must converge along the subsequence to point masses at π_ζ. Bearing this in mind, and bearing in mind that the statistical tests are defined pathwise, the previous proof is repeated, with π_* replaced by π_ζ.

Thus for every $\zeta \in \Lambda^0$, every node x that is π_ζ relevant to some player is hit infinitely often. Players' beliefs at all those nodes, and thus at the information sets they contain, converge to the empirical frequencies, which *along the subsequence* converge to π_ζ. At information sets $h \in \bar{H}(\rho_*)$, these empirical frequencies must be converging to something within $\delta/2$ of π_*, thus π_ζ is within $\delta/2$ of π_* along the path of ρ_*. Apply Lemma 8.2 to conclude that, at some date along the subsequence, some player must choose the nonexperimental portion of his strategy so that whatever the others pick, the resulting outcome is more than $\epsilon' > \epsilon$ away from ρ_*, and we have the desired contradiction. ∎

The method of proof of this proposition indicates why the result is somewhat questionable. For off-path information sets that are reached infinitely often, we cannot be sure that empirical frequencies of actions are converging to any particular limit, because calendar-time limitations on experimentation can allow capricious behavior at information sets that are visited a vanishing fraction of the time. Nonetheless along some subsequence of dates, *some* limits of empirical frequencies must be achieved at all information sets. For information sets that are reached infinitely often, our players (who are strongly asymptotically empirical) believe at those dates that these empirical frequencies are what their rivals are going to play. The statistical tests are formulated so that (asymptotically, along this subsequence of dates) all the information sets (and even nodes) relevant at the subsequence limit will be reached infinitely often, which is technically adequate for disqualifying non-Nash outcomes.

However everything depends on strong asymptotic empiricism, because this ensures that the beliefs of all players converge to a degenerate distribution at the current empirical frequencies at all information sets visited infinitely often. Under assumptions on behavior that don't guarantee convergence of behavior at off-path information sets that are visited infinitely often but with vanishing frequency, it seems somewhat silly to assume that beliefs about behavior there will converge to empirical frequencies. Unless we (and, more to the point, the players involved) have reason to believe that behavior at those information sets is converging, there is little reason to think that their beliefs will converge together with anything in particular.

9. Concluding remark

We have two reasons for regarding our results with skepticism. First, the list of assumptions we must make on behavior and beliefs is long and, to our mind, fairly unpalatable. Second, even with these assumptions, the time required for

beliefs to become approximately correct at all relevant information sets may be too long to be of practical interest. [20]

This is not to say that players cannot "learn" by some other means what they can expect at off-path or rarely encountered information sets. Perhaps on-path, frequently encountered situations provide useful data for inferring what will happen at off-path situations, if the players can make such cross-situation inferences. But a story of this sort is well beyond the models and analysis of this paper. If one trusts to the sort of argument given here in favor of Nash equilibrium (as a refinement of self-confirming equilibrium), the length of time it would take to learn what happens off the path gives further reasons for skepticism.

References

Aoyagi, M. (1992), "Evolution of beliefs and Nash equilibrium in normal form games," mimeo.

Diaconis, P., and D. Freedman (1990), "On the uniform consistency of Bayes estimates for multinomial probabilities," *The Annals of Statistics*, Vol. 18, 1317-1327.

Durrett, Richard (1991), *Probability: Theory and Examples*, Wadsworth Publishing Company, Pacific Grove, California.

Ellison, G. (1993), "Learning, local interaction, and coordination," *Econometrica*, Vol. 61, 1047–1072.

Fudenberg, D., and D. Kreps (1993), "Learning mixed equilibria," *Games and Economic Behavior*, Vol. 5, 320–367.

Fudenberg, D., and D. Kreps (1994), "Learning in extensive-form games, I: Self-confirming Equilibria," mimeo, Stanford University.

Fudenberg, D., and D. Levine (1993a), "Self-confirming equilibrium," *Econometrica*, Vol. 61, 523–546.

Fudenberg, D., and D. Levine (1993b), "Steady state learning and Nash equilibrium," *Econometrica*, Vol. 61, 547–574.

[20] Gale, Binmore, and Samuelson (1993) and Ellison (1993) make similar points in related contexts.

Gale, I., K. Binmore, and L. Samuelson (1993), "Learning to be imperfect: The ultimatum game," mimeo.

Myerson, R. (1978), "Refinements of the Nash equilibrium concept," *International Journal of Game Theory*, Vol. 7, 201–21.

Sonsino, D. (1994), "Learning to learn, pattern recognition, and Nash equilibrium," mimeo.

Appendix. Experiments taken at the level of strategies

In the MME condition, we think of experiments being taken by players information set by information set. This is quite compatible with our formulation of strategies in the behavior form, and makes for relatively easy exposition.

A different approach to the subject is possible, where we think of experiments being chosen at the level of strategies. While this formulation complicates somewhat the exposition and analysis (for reasons we will explain), it does have intuitive appeal, in that we can explicitly consider how players string together their experiments at (their own) successive information sets. We do not wish to go through all the details of this alternative approach or to give proofs, but we will sketch out here the various steps that facilitate it.

A.1. Private state variables

The expositional complication arises immediately. In order to facilitate appropriate statistical tests, we go back to the first details of the formulation of the model and imagine that each player bases his beliefs and behavior at date t on a "private state variable" ξ_t^i. That is, both $\hat{\gamma}_t^i$ and $\hat{\pi}_t^i$ takes an argument ξ_t^i. We insist that each player know at date t at least the history of past outcomes, so the (abusing notation) $\zeta_t \in \xi_t^i$ for each t. But we allow players to carry more information with them from day to day, and to use that information in formulating their actions and beliefs and (what becomes important) in conducting their statistical tests.

To have a concrete example to think of, imagine that player i determines at the start of each round of play what pure strategy s_t^i he will use that day. This pure strategy may be determined by mixing — i.e., $\hat{\pi}_t^i(\xi_t^i)$ need not be the deterministic choice of s_t^i — but in this case we imagine that i does his mixing at the outset, with s_t^i the result. Then we can think of $\xi_t^i = (z_1, z_2, \ldots, z_{t-1}, s_1^i, s_2^i, \ldots, s_{t-1}^i) = (\zeta_t, s_1^i, s_2^i, \ldots, s_{t-1}^i)$; i.e., i remembers past outcomes and the pure strategies he meant to employ at each date.

Still in terms of formulation, let ξ^i denote a typical infinite (private) history for player i, which contains ζ. Let Ξ^i denote the space of all ξ^i. We write $\xi = (\xi^1, \ldots, \xi^I)$ for an array of private histories, and we say that ξ is *consistent* if all information in ξ^i that is common to information in ξ^j (such as the outcome in each period) is consistent with the information in ξ^j. Then $\Xi \subset \prod_i \Xi^i$ denotes the set of *consistent* private history vectors. We also use ξ_t to denote an array of partial private histories — i.e., $\xi_t = (\xi_t^1, \ldots, \xi_t^I)$ — and Ξ_t denotes the set of consistent partial private history arrays.

Since $\zeta_t \in \xi_t^i$ for each i and t, only cosmetic changes are needed for the formulations of strong asymptotic empiricism and asymptotic myopia (of either variety). The definitions of unstability and weak stability are also changed only cosmetically, where we work with a probability measure $\mathbf{P}$ defined on Ξ, which is assumed to have the appropriate marginals on $\mathcal{Z}$ (onto which $\xi \in \Xi$ can be projected) for the given behavior rules.

In this richer setting, it is relatively easy to see that none of our earlier results are changed. It didn't matter that players have access to private information, since (under the terms of stability) they asymptotically ignore any such information in formulating their behavior.

Having made the assertion in the preceding paragraph, a caveat is in order. We do not preclude the possibility that two players i and j share private information that k does not have. This permits i and j to act in correlated fashion,

relative to what k observes. Under appropriate conditions, we might in consequence find that play converges to a correlated equilibrium of the underlying game. But insofar as behavior is asymptotically stationary (independent of any extraneous information), all our results apply. That is, our results remain valid — roughly, if there is convergence to an asymptotically stationary state, that state must be a confirmed expectations/Nash equilibrium (depending on the level of experimentation, etc.) — but because it is more plausible that players will correlate using partially private information, it is less likely that the antecedent clause — there is convergence... — will pertain.

A.2. The infinite strategic-experimentation condition

Fix player i and node $x \in X$ such that $h(x) \notin H^i$. Define $S^i(x)$ to be the set of all pure strategies for player i such that if i plays any $s^i \in S^i(x)$, then i does not preclude hitting x.[21]

Definition. *The behavior rule $\hat{\pi}^i$ satisfies the* ***infinite strategic-experimentation condition*** *if for every $\xi \in \Xi$ and x such that $h(x) \notin H^i$,*

$$\sum_{n=1}^{\infty} \hat{\pi}_t^i(\xi_t^i)(S^i(x)) = \infty, \tag{A.1}$$

where ξ_t^i is the date t, player i subhistory of ξ and $\hat{\pi}_t^i(\xi_t^i)(S^i(x))$ is the probability that the behaviorally mixed strategy $\hat{\pi}_t^i(\xi_t^i)$ selects some pure strategy from $S^i(x)$.[22]

The formal relevance of this definition is established by the following proposition.

Proposition A.1. *Fix a strategy profile π_* and a player i. Suppose that i's behavior rule satisfies the infinite strategic-experimentation condition and, for some $\Lambda \in \Xi$, there*

[21] In other words, for every $x' \prec x$ such that $h(x') \in H^i$, $s^i(h(x'))$ prescribes the unique action that leads from x' towards x, for $s^i \in S^i(x)$.

[22] Let $B(x)$ be the set of actions that lead to the node x, and let $B^i(x) = B(x) \cap A^i$. Then $\hat{\pi}_t^i(\xi_t^i)(S^i(x)) = \prod_{a \in B^i(x)} \hat{\pi}_t^i(\xi_t^i)(a)$.

exists $\epsilon > 0$ *such that for all* $\xi \in \Xi$ *and for all* $j \neq i$, $\hat{\pi}_t^j(\xi_t^j)(a) > \epsilon$ *for all* $a \in A^j$ *such that* $\pi_*^j(a) > 0$. *Then every information set* h *that is relevant to* i *at* π_* *will be hit infinitely often, almost surely on* Ξ.

This proposition is a quick corollary to the following generalization of the Borel–Cantelli Lemma.

Lemma A.1. *For any probability space* $(\Omega, F, \mathbf{P})$, *sequence of increasing sub-*σ*-fields* $\{F_t\}$, *and sequence of events* $\{A_t\}$, *with* $A_t \in F_t$ *for each* t,

$$\left\{\omega \ : \ \omega \in A_t \text{ for infinitely many } t\right\} = \left\{\omega \ : \ \sum_{t=1}^{\infty} \mathbf{P}(A_{t+1}|F_t) = \infty.\right\}$$

The equality between the sets in the lemma is, of course, up to a $\mathbf{P}$-null set. This lemma is easily proved using martingale theory; see, e.g., Durrett (1991, p.208).

Several comments about the definition and result are in order:

(1) It may be helpful to describe two different ways that condition (A.1) might be satisfied.[23] First, suppose that for some sequence of nonnegative numbers $\{\delta_t\}$ such that $\sum_{t=1}^{\infty} \delta_t = \infty$ and $\lim_{t\to\infty} \delta_t = 0$, we have $\hat{\pi}_t^i(\xi_t^i)(S^i(x)) \geq \delta_t$ uniformly in x. That is, player i experiments with different strategies at random, at rates which vanish as calendar time passes, but sufficiently slowly so that enough experiments are taken. We will refer to this way of satisfying (A.1) *slowly-vanishing random experimentation*. Note that slowly-vanishing random experimentation will pose no problems for asymptotic myopia regardless of limitations put on experiments, because as long as the bounds on the probability of experimentation go to zero faster than the sequence $\{\epsilon_t\}$ used for asymptotic myopia, slowly-vanishing random experimentation can be accomodated within the "nonexperimental" portion of the player's behavior rule.

[23] These are each special cases. Condition (A.1) is a good deal more general than either.

A second way (A.1) is satisfied is if players experiment consciously with non-vanishing probability, but at a frequency whose proportion falls to zero. Specifically, we might have, for each x, an infinite sequence of dates $t_1(x), t_2(x), \ldots$ and a $\beta > 0$ such that $\hat{\pi}^i_{t_k}(\xi^i_{t_k})(S^i(x)) \geq \beta$ for all k and $k/t_k(x) \to 0$. We call this way of satisfying (A.1) *specific-date experimentation.* Note that the condition $k/t_k(x) \to 0$ is imposed to make this compatible with calendar-time limitations on experiments, but there may remain problems for experience-time limitations.

(2) As the proof of Proposition A.1 makes clear, (A.1) ensures that (under appropriate conditions on the behavior of i's rivals) that every *node* x that is relevant to i will be reached infinitely often. Our aim is to ensure (only) that every relevant *information set* is reached infinitely often, and thus requiring (A.1) for every node x is stronger than necessary. Moreover, examples show that requiring (A.1) for every x may be unreasonably restrictive; see our earlier discussion (on pp.21ff) concerning different experiments that lead to the same information set.

(3) Bearing in mind point (2) immediately above — the objective of reaching every relevant node infinitely often may be more than we want — if we pursue that objective, then the lemma implies that (A.1) is necessary: For a given node x, x will not be reached infinitely often (almost surely), no matter what i's rivals do, unless i's behavior rule satisfies (A.1) for x.

(4) A strength of the MME condition is that players are not required to experiment at information sets that are seemingly irrelevant; cf. the discussion surrounding Figure 1 (page 23). Something roughly similar goes on here, although it is far from transparent: Condition (A.1) gives player i credit for the intention to experiment at information set h at date t (even randomly), even if h is not reached. Thus (to give a very concrete example) if Player 1 chooses A_1 at date t with probability $1/t$, and Player 2 chooses A_2 at date t with probability $1/t$, then Player 2 will

have fulfilled her obligations under (A.1) and yet will actually choose A_2 only finitely often, almost surely.

(5) As with MME or any lower bound on information-set-based experimentation that implies infinite experimentation almost surely, the definition asks for more experimentation than would be optimal for rational (discounted payoff) play against a multi-armed bandit.

A.3. Statistical tests and results

From Proposition A.1 to the sorts of results we pursue is straightforward *if* behavior meets the extra condition of Proposition A.1; roughly, that behavior is uniformly nonexperimental. Having shifted from MME to strategic experimentation has not changed our lack of enthusiasm for the uniform nonexperimentation condition, nor does it moot the problems raised by the example in Figure 2. Thus to get satisfactory results with minimum strategic experimentation, we must again enlist statistical test sequences.

This is where the extra baggage of personal histories becomes useful. Previously we suggested as an example that each player remember the pure strategy he would have employed at each date; now we insist that each player include this in his personal history, together with the history of outcomes. This allows the formulation of a simple statistical test sequence that is adequate in this setting: We begin with notation. For each $x \in X$, let $w(x)$ denote the initial node that precedes x, let $B(x)$ be the set of actions that leads to x, and let $\kappa^i(S^i(x); \xi_t^i)$ denote the number of times along the personal history ξ_t^i that i chose (a priori) a strategy $s^i \in S^i(x)$. At date t, for each information set $x \in X$ such that $h(x) \notin H^i$, and for each strategy profile π^{-i} for i's rivals, we imagine that i computes the number of times that x "should have" been reached if his rivals play according to π^{-i} in i.i.d. fashion:

$$\mu^i(x, \pi^{-i}, \xi_t^i) = \rho(w(x)) \times \prod_{a \in B(x) \cap A^{-i}} \pi^{-i}(a) \times \kappa^i(S^i(x); \xi_t^i).$$

Then the test is: For some nondecreasing sequence $\{\eta_k\}$ with limit infinity and for some $\gamma_* > 0$,

$$\xi_t^i \notin \Lambda_t^i \qquad \text{if} \qquad \hat{\gamma}_t^i(\xi_t^i)\{\pi^{-i} : \kappa(x;\xi_t^i) < \eta_{\mu^i(x,\pi^{-i},\xi_t^i)} \text{ for some } x\} > \gamma^*.$$

In words (and a bit roughly), if i is attaching significant probability to π^{-i} such that, for some x, the number of visits to x significantly less than what can be expected, than i rejects. Note that *what can be expected* here is very weak; all that we insist upon is that we *expect* a number of visits which goes to infinity in the "expectation" given by μ^i. That is, in the spirit of our earlier statistical tests we might think to insist that $\eta_k \sim k$. But in fact, even $\eta_k = \ln\ln\ln k$ will do.

The point of this test should be clear. If i believes (asymptotically) that node x is relevant, and if he satisfies the infinite strategic-experimentation condition, then μ^i must go to infinity, and x must be visited infinitely often. Using the sort of induction argument employed in the proofs of Propositions 7.1 and 8.1, we find that every relevant node will be reached infinitely often, and thus non-Nash profiles/outcomes are unstable.